

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-1

Perfect score: 25

Sequence: 1 aaaaaaaacacgagccagcatgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_to.\*

10: gb\_sts.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_mtg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	221399	9	AL604045 Mouse DNA
2	23.4	93.6	49499	8	AF184110 Homo sapi
3	23.4	93.6	51905	6	AR659626 Sequence
4	23.4	93.6	51905	6	AR659740 Sequence
5	23.4	93.6	61412	14	AC021493 Homo sapi
6	23.4	93.6	139280	8	AC099052 Homo sapi
7	23.4	93.6	142943	8	AC099405 Homo sapi
8	23.4	93.6	150221	8	AC004593 Homo sapi
9	23.4	93.6	163473	14	AC144870 Human DNA
10	23.4	93.6	165802	8	AL135914 Homo sapi
11	23.4	93.6	168813	8	AP000555 Homo sapi
12	23.4	93.6	171272	8	AC146225 Pan trogl
13	23.4	93.6	171461	8	AC092041 Homo sapi
14	23.4	93.6	184515	8	AC010653 Homo sapi
15	23.4	93.6	185161	8	AC006059 Homo sapi
16	23.4	93.6	187530	14	AC096875 Pan trogl
17	23.4	93.6	199471	8	AC009097 Homo sapi
18	23.4	93.6	328439	14	AL590137 Homo sapi

AB030316	Mus muscu	9	AB030316	12593	89.6	22.4	C 19
BV02806	RPANMSEQ	10	BV02806	394	87.2	21.8	C 20
BV09496	RPANMSEQ	10	BV09496	398	87.2	21.8	C 21
BV159481	RPANMSEQ	10	BV159481	398	87.2	21.8	C 22
AF270524	Homo sapi	8	AF270524	701	87.2	21.8	C 23
BV69872	S21696216	10	BV69872	749	87.2	21.8	C 24
BV590677	G5916435	10	BV590677	861	87.2	21.8	C 25
BV594288	bxj38a02.	10	BV594288	942	87.2	21.8	C 26
AK090515	Homo sapi	8	AK090515	2141	87.2	21.8	C 27
AK097083	Homo sapi	8	AK097083	2530	87.2	21.8	C 28
S81092	acyl-coenzy	9	S81092	3041	87.2	21.8	C 29
L42233	Mus musculu	9	MUSACACT	3697	87.2	21.8	C 30
BS571793	Human DNA	8	BS571793	35616	87.2	21.8	C 31
AC104230	Homo sapi	14	AC104230	46784	87.2	21.8	C 32
AL358393	Human DNA	8	AL358393	50175	87.2	21.8	C 33
AC104800	Homo sapi	8	AC104800	54618	87.2	21.8	C 34
AC105044	Homo sapi	14	AC105044	65882	87.2	21.8	C 35
AL589910	Human DNA	8	AL589910	66591	87.2	21.8	C 36
AC133141	Homo sapi	8	AC133141	70560	87.2	21.8	C 37
AL591766	Human DNA	8	AL591766	73084	87.2	21.8	C 38
AP000280	Homo sapi	8	AP000280	73687	87.2	21.8	C 39
AC026374	Homo sapi	8	AC026374	75385	87.2	21.8	C 40
AL096775	Human DNA	8	HSDJ336M4	76241	87.2	21.8	C 41
AP000039	Homo sapi	8	AP000039	100000	87.2	21.8	C 42
AP000107	Homo sapi	8	AP000107	100000	87.2	21.8	C 43
AP000183	Homo sapi	8	AP000183	100000	87.2	21.8	C 44
Continuation (2 of		14	AC145312_1	110000	87.2	21.8	C 45
AC012356	Homo sapi	8	AC012356	112527	87.2	21.8	C 46
AL109811	Human DNA	8	HSJ635E18	112769	87.2	21.8	C 47
AL008721	Human DNA	8	HSJ390C10	114231	87.2	21.8	C 48
AC078901	Homo sapi	14	AC078901	115909	87.2	21.8	C 49
AL121820	Human chr	8	CNS01DSM	117449	87.2	21.8	C 50
AP000609	Homo sapi	8	AP000609	130964	87.2	21.8	C 51
AC008131	Homo sapi	8	AC008131	131891	87.2	21.8	C 52
AC006001	Homo sapi	8	AC006001	135044	87.2	21.8	C 53
AL954211	Pan trogl	8	AL954211	138573	87.2	21.8	C 54
AC103718	Homo sapi	8	AC103718	144801	87.2	21.8	C 55
AC146516	Homo sapi	14	AC146516	147478	87.2	21.8	C 56
AJ300188	Homo sapi	8	HSJ3000188	148049	87.2	21.8	C 57
AC044819	Homo sapi	14	AC044819	151798	87.2	21.8	C 58
AC011921	Homo sapi	8	AC011921	153922	87.2	21.8	C 59
AC148077	Pan trogl	14	AC148077	155690	87.2	21.8	C 60
AC092198	Homo sapi	8	AC092198	158456	87.2	21.8	C 61
AC166354	Mus muscu	14	AC166354	160029	87.2	21.8	C 62
AC120013	Mus muscu	14	AC120013	161646	87.2	21.8	C 63
AC087434	Pan trogl	8	AC087434	161716	87.2	21.8	C 64
AC145756	Pan trogl	14	AC145756	162240	87.2	21.8	C 65
AC023421	Homo sapi	8	AC023421	163365	87.2	21.8	C 66
AC109460	Homo sapi	8	AC109460	163424	87.2	21.8	C 67
AL391259	Human DNA	8	AL391259	163520	87.2	21.8	C 68
AC010823	Homo sapi	14	AC010823	164376	87.2	21.8	C 69
AC068023	Homo sapi	14	AC068023	164383	87.2	21.8	C 70
AC161221	Mus muscu	14	AC161221	164547	87.2	21.8	C 71
AC067860	Homo sapi	14	AC067860	165264	87.2	21.8	C 72
AL45467	Human DNA	8	AL45467	166336	87.2	21.8	C 73
AL4591722	Homo sapi	14	AL4591722	166439	87.2	21.8	C 74
AL356864	Homo sapi	14	AL356864	166876	87.2	21.8	C 75
AC101652	Mus muscu	9	AC101652	168183	87.2	21.8	C 76
AC079047	Homo sapi	14	AC079047	169372	87.2	21.8	C 77
AC138627	Homo sapi	8	AC138627	170920	87.2	21.8	C 78
AC016178	Homo sapi	8	AC016178	171035	87.2	21.8	C 79
AL355517	Human DNA	8	AL355517	171348	87.2	21.8	C 80
AC065943	Homo sapi	14	AC065943	171544	87.2	21.8	C 81
AL160396	Human DNA	8	AL160396	172437	87.2	21.8	C 82
AC140275	Mus muscu	9	AC140275	173022	87.2	21.8	C 83
AL590292	Homo sapi	14	AL590292	173751	87.2	21.8	C 84
AF216798	Homo sapi	14	AF216798	174295	87.2	21.8	C 85
AC135506	Homo sapi	8	AC135506	174363	87.2	21.8	C 86
AC069125	Homo sapi	14	AC069125	176010	87.2	21.8	C 87
AL353807	Human DNA	8	AL353807	179064	87.2	21.8	C 88
AL314486	Homo sapi	8	AL314486	179070	87.2	21.8	C 89
AC149106	Papio anu	14	AC149106	179560	87.2	21.8	C 90
AL355840	Human chr	8	CNS05TCV	179937	87.2	21.8	C 91



92	21.8	87.2	181842	9	AC137127	AC137127 Mus muscu
93	21.8	87.2	183483	14	AC136510	AC136510 Pan trogl
94	21.8	87.2	183539	8	AC092268	AC092268 Homo sapi
95	21.8	87.2	184000	8	AP002449	AP002449 Homo sapi
96	21.8	87.2	184295	14	AC021138	AC021138 Homo sapi
97	21.8	87.2	185146	14	AC104311	AC104311 Homo sapi
98	21.8	87.2	186134	9	AL591390	AL591390 Mouse DNA
99	21.8	87.2	187788	14	AC138925	AC138925 Homo sapi
100	21.8	87.2	191137	14	AC145064	AC145064 Pan trogl
101	21.8	87.2	191196	14	AL441903	AL441903 Homo sapi
102	21.8	87.2	191957	8	AC015845	AC015845 Homo sapi
103	21.8	87.2	194160	8	AC130472	AC130472 Homo sapi
104	21.8	87.2	194464	14	AC146119	AC146119 Pan trogl
105	21.8	87.2	195558	8	AC007448	AC007448 Homo sapi
106	21.8	87.2	196954	8	AC008267	AC008267 Homo sapi
107	21.8	87.2	197683	9	AC118476	AC118476 Mus muscu
108	21.8	87.2	200028	8	AC092680	AC092680 Homo sapi
109	21.8	87.2	201020	8	AC022809	AC022809 Homo sapi
110	21.8	87.2	201709	14	AC145497	AC145497 Papio anu
111	21.8	87.2	207860	14	AP001907	AP001907 Homo sapi
112	21.8	87.2	209472	14	AC090759	AC090759 Homo sapi
113	21.8	87.2	214019	14	AC064803	AC064803 Mus muscu
114	21.8	87.2	214331	14	AC145462	AC145462 Callithri
115	21.8	87.2	219646	14	AC139247	AC139247 Homo sapi
116	21.8	87.2	219646	14	AC139247	AC139247 Homo sapi
117	21.8	87.2	228363	9	AC116128	AC116128 Mus muscu
118	21.8	87.2	232188	9	AL672055	AL672055 Mouse DNA
119	21.8	87.2	232401	14	AC106137	AC106137 Rattus no
120	21.8	87.2	240058	8	AC110921	AC110921 Homo sapi
121	21.8	87.2	247451	14	AC097417	AC097417 Rattus no
122	21.8	87.2	248281	8	AC008737	AC008737 Homo sapi
123	21.8	87.2	256019	8	AC016525	AC016525 Homo sapi
124	21.8	87.2	259474	8	HUAC004605	HUAC004605 Homo sapi
125	21.8	87.2	340000	8	AP001715	AP001715 Homo sapi
126	21.4	85.6	215965	14	AC151869	AC151869 Callithri
127	21	84.0	204793	8	AC026333	AC026333 Homo sapi
128	20.8	83.2	401	10	BV190162	BV190162 sqmml6629
129	20.8	83.2	55673	8	AC117944	AC117944 Homo sapi
130	20.8	83.2	96322	8	AL512640	AL512640 Human DNA
131	20.8	83.2	97832	8	HSJ46801	HSJ46801 Human DNA
132	20.8	83.2	101324	8	AC010386	AC010386 Homo sapi
133	20.8	83.2	104947	8	AC008116	AC008116 Homo sapi
134	20.8	83.2	107469	8	HS29H4	AL078475 Homo sapi
135	20.8	83.2	112895	8	AC107210	AC107210 Homo sapi
136	20.8	83.2	117879	8	AC110004	AC110004 Homo sapi
137	20.8	83.2	129090	14	AC083961	AC083961 Homo sapi
138	20.8	83.2	132070	8	AC003663	AC003663 Homo sapi
139	20.8	83.2	146947	14	AC012515	AC012515 Homo sapi
140	20.8	83.2	150810	8	AC018500	AC018500 Homo sapi
141	20.8	83.2	152333	8	AC010476	AC010476 Homo sapi
142	20.8	83.2	157356	8	AL590783	AL590783 Human DNA
143	20.8	83.2	160298	14	AC067818	AC067818 Homo sapi
144	20.8	83.2	161139	8	AC132812	AC132812 Homo sapi
145	20.8	83.2	162364	8	AP005213	AP005213 Homo sapi
146	20.8	83.2	163489	8	AC016697	AC016697 Homo sapi
147	20.8	83.2	164888	14	AC024101	AC024101 Homo sapi
148	20.8	83.2	165919	8	AC026463	AC026463 Homo sapi
149	20.8	83.2	176600	14	AC023546	AC023546 Homo sapi
150	20.8	83.2	178050	8	AC142324	AC142324 Pan trogl

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1

Whitehead,S.

Direct Submission

Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Apr 7, 2002 this sequence version replaced gi:17127811. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep RP23-418011 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers

1..221399

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="11"

/clone="RP23-418011"

/clone\_lib="RPCI-23"

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 221399;

Best Local Similarity 100.0%; Pred. No. 0.4; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;

QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25

|||||

Db 21788 AAAAAAAAAATCAGCGCCAGCATGG 21812

RESULT 2

AF184110/c

LOCUS AF184110 49499 bp DNA linear PRI 26-SEP-1999

DEFINITION Homo sapiens cyclophilin-related protein (NKTR) gene, complete cds.

ACCESSION AF184110

VERSION AF184110.1 GI:5923890

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 49499)

AUTHORS Anderson,S.K., Gallinger,S., Roder,J., Frey,J., Young,H.A. and Ortaldo,J.R.

TITLE A cyclophilin-related protein involved in the function of natural killer cells

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (2), 542-546 (1993)

PUBMED 8421688

REFERENCE 2 (bases 1 to 49499)

AUTHORS Anderson,S.K.

TITLE Structure of the human NKTR gene  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 49499)  
AUTHORS Anderson, S.K.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1999) Lab of Expt. Immunology, NCI-FCRDC,  
Building 560, Room 31-93, Frederick, MD 21702-1201, USA  
FEATURES  
source 1..49499  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
map="3p21-p23"  
1300..49223  
/gene="NKTR"  
join(1300..1464,1670..1750,18258..18332,19708..19815,  
20352..20396,22117..22204,31234..31263,31859..32004,33289..33289,  
33289..33511,34268..34423,35301..35388,35909..36054,  
37556..40438,43189..43284,44027..44083,44590..44691,  
46576..49223)  
/gene="NKTR"  
join(1693..1750,18258..18332,19708..19815,20352..20396,  
22117..22204,31234..31263,31859..32004,33289..33511,  
34268..34423,35301..35388,35909..36054,37556..40438,  
43189..43284,44027..44083,44590..44691,46576..46663)  
/gene="NKTR"  
/note="RS-cyclophilin"  
/codon\_start=1  
/product="cyclophilin-related protein"  
/protein\_id="AAB56402.1"  
/db\_xref="GI:5923891"  
/translation="MQAQRPOCHFDIEINREPVGRIMFQFSDICPKTKNFCILCS  
HDEGLTKGKCYGSGTFRVVKVNFMIQGDPSFEGNGKGSIEYGGYKDFENFLIK  
HDFALFSMANRGHTNGSQFFITTPAPHLDDGVHVFGLVIGFEVIEQIENLKTD  
ASRYADRVVDCVLTATKGIKDFEKKRKPHTSGSDSSNSSESSESELEH  
ESRRRKHRRPKVKRRKEASSSEEPKRNHAMNPKGHSERSDTNEKRSVDSAK  
REKVPVPEIPVPEPFRLLRRDMPVVTAEPEPKI PDVAPIVSDQKPSVSKGRKIK  
GRGIRVHTPPSRSCSDDSDSETPPHWKEEMQRLRAYRPPSGEWSKDGKLDSP  
CSSRWDRSQRSSRSYNGYSDLSFTPHKRRKKRKKVKKKKKKKKKKR  
HKQTKRRIILPDSIESKSTRMSKCDRERSRSSLSHSHSKDWSKDQVQ  
SSLTHSRDSYRSKSHSQYSGRSRSTASKSHSRSSRSKSSSKSHRKRASKS  
PKRTASOLSENKPVKTRPLRATMAQENNVVQVPAENIPVPLSDPPSPRWKPGQK  
PKWSYRIQEMAKTTHLLPIQSTYSLANIKETGSSSVYHKEKNSDQSTVSKYS  
DSSSPSRSSRSRSHRSYRSLASSRSHRSRSPSRSHRNKYSKSDSQCSR  
SSYTSISDGRRAKRLSSGKNSVSHKSSSEKTLHXYKVRDRSSCVRY  
SESRSLDSDSQSSVQATQSAKEKQGMERTHKNQKNGKEKSEKSECEPHS  
KRTLKNSLPHLRGSKPKRKNYAGSKWDSERDVTNKNKNDSPHSKDEEGEA  
TSDSESVSEHIIKVPKTKSTSLPDDNGAWKSKORTSTSDSGSCSNENNRG  
KPQKHGSKENLAKRETKVKEKLGKKKKKAPKPKQAFHWQPLERGEEREEI  
DDQVTKSEKVKYSNNETIKDNLKTEKSEEDLSGKHDVTVSSDLOQTKDDSK  
LSISPTALNTEENACIQNIQHVSEVPNGVEDVLTDDNMEICTPDRSPAPKYEET  
PLGNARLDTPDNI VLQDMATEHPQAEVVKQESMESKVLGEVQKDSQSLASA  
GESGKKEVAESKQINLIDKKKPLQGVNLAA PNAATSSAVEVKVLTTPMKPQGL  
RIEIKSNKVRPGSLFVEVKTARLNRRPNQESSDEQTPSRDDDSQSRSPSRSK  
SETKRRHTRSYSHRSRSGRSTSTSRYSRSGRSGRSGRSGRSGRSGRSGRSGR  
SHRTSRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGRSGR  
RSYGDSGSDRSYSHRSRSPSRSSRY"  
polyA\_signal 49203..49208  
/gene="NKTR"  
polyA\_site 49223  
/gene="NKTR"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 8; Length 49499;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 16079 AAAAAAAAAATCAGCGCCAGGCATGG 16055

## RESULT 3

AR659626/c  
LOCUS AR659626 51905 bp DNA linear PAT 13-JUN-2005  
DEFINITION Sequence 667 from patent US 6900016.  
ACCESSION AR659626  
VERSION AR659626.1 GI:67595666  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

## REFERENCE

1 (bases 1 to 51905)  
Venter, J.C., Zhang, J.N., Liu, X., Rowe, W., Cravchik, A., Kalush, F.,  
Naik, A., Subramanian, G. and Woodage, T.  
Polymorphisms in known genes associated with inflammatory  
autoimmune disease, methods of detection and uses thereof  
Patent: US 6900016-A 667 31-MAY-2005;  
Applera Corporation; Norwalk, CT

## FEATURES

source 1..51905  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 6; Length 51905;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 16760 AAAAAAAAAATCAGCGCCAGGCATGG 16736

## RESULT 4

AR659740/c  
LOCUS AR659740 51905 bp DNA linear PAT 13-JUN-2005  
DEFINITION Sequence 781 from patent US 6900016.  
ACCESSION AR659740  
VERSION AR659740.1 GI:67595794  
KEYWORDS  
SOURCE Unknown.

## REFERENCE

1 (bases 1 to 51905)  
Venter, J.C., Zhang, J.N., Liu, X., Rowe, W., Cravchik, A., Kalush, F.,  
Naik, A., Subramanian, G. and Woodage, T.  
Polymorphisms in known genes associated with inflammatory  
autoimmune disease, methods of detection and uses thereof  
Patent: US 6900016-A 781 31-MAY-2005;  
Applera Corporation; Norwalk, CT

## FEATURES

source 1..51905  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 6; Length 51905;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 16760 AAAAAAAAAATCAGCGCCAGGCATGG 16736

## RESULT 5

AC021493/c  
LOCUS AC021493 61412 bp DNA linear HTG 13-JUL-2000  
DEFINITION Homo sapiens clone RP11-104F3, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC021493  
VERSION AC021493.2 GI:9146785  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

1 (bases 1 to 61412)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## JOURNAL

Homo sapiens, clone RP11-104F3

## REFERENCE

2 (bases 1 to 61412)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Chopel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gace,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Jul 13, 2000 this sequence version replaced gi:6705607.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5374

Center clone name: 104\_F\_3

-----

\* NOTE: This record contains 63 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 896: contig of 896 bp in length  
\* 897 996: gap of 100 bp  
\* 997 1867: contig of 871 bp in length  
\* 1868 1967: gap of 100 bp  
\* 1968 2874: contig of 907 bp in length  
\* 2875 2974: gap of 100 bp  
\* 2975 3849: contig of 875 bp in length  
\* 3850 3949: gap of 100 bp  
\* 3950 4841: contig of 892 bp in length  
\* 4842 4941: gap of 100 bp  
\* 4942 5805: contig of 864 bp in length  
\* 5806 5905: gap of 100 bp  
\* 5906 6772: contig of 867 bp in length  
\* 6773 6872: gap of 100 bp  
\* 6873 7746: contig of 874 bp in length  
\* 7747 7846: gap of 100 bp  
\* 7847 8733: contig of 887 bp in length  
\* 8734 8833: gap of 100 bp  
\* 8834 9702: contig of 869 bp in length  
\* 9703 9802: gap of 100 bp  
\* 9803 10666: contig of 864 bp in length

\* 10667 10766: gap of 100 bp  
\* 10767 11621: contig of 855 bp in length  
\* 11622 11721: gap of 100 bp  
\* 11722 12583: contig of 862 bp in length  
\* 12584 12683: gap of 100 bp  
\* 12684 13559: contig of 876 bp in length  
\* 13560 13659: gap of 100 bp  
\* 13660 14537: contig of 878 bp in length  
\* 14538 14637: gap of 100 bp  
\* 14638 15223: contig of 886 bp in length  
\* 15224 15624: gap of 100 bp  
\* 15624 16568: contig of 845 bp in length  
\* 16469 17437: contig of 869 bp in length  
\* 16569 17537: gap of 100 bp  
\* 17438 18397: contig of 860 bp in length  
\* 17538 18497: gap of 100 bp  
\* 18398 19355: contig of 858 bp in length  
\* 18498 19455: gap of 100 bp  
\* 19356 20366: contig of 911 bp in length  
\* 19456 20466: gap of 100 bp  
\* 20367 21361: contig of 895 bp in length  
\* 20467 21461: gap of 100 bp  
\* 21362 22361: contig of 900 bp in length  
\* 21462 22461: gap of 100 bp  
\* 22362 23361: contig of 900 bp in length  
\* 22462 23461: gap of 100 bp  
\* 23362 24327: contig of 866 bp in length  
\* 23462 24428: gap of 100 bp  
\* 24328 25319: contig of 892 bp in length  
\* 25320 25419: gap of 100 bp  
\* 25420 26258: contig of 839 bp in length  
\* 26259 26358: gap of 100 bp  
\* 26359 27250: contig of 892 bp in length  
\* 27251 27351: gap of 100 bp  
\* 27352 28229: contig of 879 bp in length  
\* 28230 28320: gap of 100 bp  
\* 28321 29174: contig of 845 bp in length  
\* 29175 29274: gap of 100 bp  
\* 29275 30135: contig of 861 bp in length  
\* 30136 30235: gap of 100 bp  
\* 30236 31099: contig of 864 bp in length  
\* 31100 31199: gap of 100 bp  
\* 31200 32108: contig of 909 bp in length  
\* 32109 32208: gap of 100 bp  
\* 32209 33061: contig of 853 bp in length  
\* 33062 33161: gap of 100 bp  
\* 33162 34022: contig of 861 bp in length  
\* 34023 34122: gap of 100 bp  
\* 34123 34973: contig of 851 bp in length  
\* 34974 35073: gap of 100 bp  
\* 35074 35947: contig of 874 bp in length  
\* 35948 36047: gap of 100 bp  
\* 36048 36920: contig of 873 bp in length  
\* 36921 37020: gap of 100 bp  
\* 37021 37909: contig of 889 bp in length  
\* 37910 38009: gap of 100 bp  
\* 38010 38896: contig of 887 bp in length  
\* 38897 38996: gap of 100 bp  
\* 38997 39872: contig of 876 bp in length  
\* 39873 39972: gap of 100 bp  
\* 39973 40874: contig of 902 bp in length  
\* 40875 40974: gap of 100 bp  
\* 40975 41836: contig of 862 bp in length  
\* 41837 41936: gap of 100 bp  
\* 41937 42810: contig of 874 bp in length  
\* 42811 42910: gap of 100 bp  
\* 42911 43793: contig of 883 bp in length  
\* 43794 43893: gap of 100 bp  
\* 43894 44796: contig of 903 bp in length  
\* 44797 44896: gap of 100 bp  
\* 44897 45751: contig of 855 bp in length  
\* 45752 45851: gap of 100 bp

JOURNAL REFERENCE  
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
REFERENCE 3 (bases 1 to 139280)  
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (27-APR-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
COMMENT On Apr 27, 2002 this sequence version replaced gi:16799010.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgchgs@u.washington.edu  
Drafting Center: BCM  
----- Project Information  
Center project name: chr-3  
Center Clone name: RP11-437M11 (bc0424)  
----- Summary Statistics  
Sequencing vector: unknown; 33% of reads  
Chemistry: Dye-terminator ET; 76% of reads  
Chemistry: Dye-terminator Big Dye; 24% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 139214 bases at least Q40  
Consensus quality: 139263 bases at least Q30  
Insert size: 139280; sum-of-contigs  
Quality coverage: 13.0x in Q20 bases; sum-of-contigs  
-----  
Overlapping Sequences:  
5': RP11-219121 (UWGC:bc0309) AC092047, 130783-bp overlap  
3': RP4-613B23 AC006059, 101219-bp clone overlap, of which  
37845 bp is included in this submission  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an  
alternate chemistry or covered by high quality data (i.e., phred  
quality >= 30); an attempt was made to resolve all sequencing  
problems, such as compressions and repeats; all regions were  
covered by at least one plasmid subclone or more than one M13  
subclone; and the assembly was confirmed by restriction digest.  
-----  
Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
fingerprinting. Comparison of the experimentally derived digest  
fragments with sequence-predicted fragments is given below.  
The electronically-digested sequence consists of both insert and  
vector, in order to accurately represent the entire circular BAC.  
Small fragments below a variable cutoff (approximately 400-800 bp)  
are not resolved in the fingerprint and hence do not appear  
in the table. There are no significant remaining discrepancies  
between the experimental and predicted values. Uniquely ordered  
fragments are separated by dashed lines.  
-----  
Bg11 EcoRI HindIII  
SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt  
-----  
-----  
-----

45852 46735: contig of 884 bp in length  
\* 46835: gap of 100 bp  
\* 46836 47737: contig of 902 bp in length  
\* 47738 47837: gap of 100 bp  
\* 47838 48701: contig of 864 bp in length  
\* 48702 48801: gap of 100 bp  
\* 48802 49713: contig of 912 bp in length  
\* 49714 49813: gap of 100 bp  
\* 49814 50709: contig of 896 bp in length  
\* 50710 50809: gap of 100 bp  
\* 50810 51714: contig of 905 bp in length  
\* 51715 51814: gap of 100 bp  
\* 51815 52714: contig of 900 bp in length  
\* 52715 52814: gap of 100 bp  
\* 52815 53673: contig of 859 bp in length  
\* 53674 53773: gap of 100 bp  
\* 53774 54631: contig of 858 bp in length  
\* 54632 54731: gap of 100 bp  
\* 54732 55624: contig of 893 bp in length  
\* 55625 55724: gap of 100 bp  
\* 55725 56572: contig of 848 bp in length  
\* 56573 56672: gap of 100 bp  
\* 56673 57559: contig of 887 bp in length  
\* 57560 57659: gap of 100 bp  
\* 57660 58551: contig of 892 bp in length  
\* 58552 58651: gap of 100 bp  
\* 58652 59496: contig of 845 bp in length  
\* 59497 59596: gap of 100 bp  
\* 59597 60448: contig of 852 bp in length  
\* 60449 60549: gap of 100 bp  
\* 60549 61412: contig of 864 bp in length.  
-----  
Location/Qualifiers  
1..61412  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-104F3"  
/clone\_lib="RPC1-11 Human Male BAC"  
897..996  
/estimated\_length=100  
1868..1967  
/estimated\_length=100  
2875..2974  
/estimated\_length=100  
3850..3949  
/estimated\_length=100  
4842..4941  
-----  
Query Match 93.6%; Score 23.4; DB 14; Length 61412;  
Best Local Similarity 96.08; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
-----  
Qy 1 AAAAAAAAAATCAGCCGAGCATGG 25  
|||||  
Db 58155 AAAAAAAAAATCGCCGAGCATGG 58131  
|||||  
-----  
RESULT 6  
AC099052/c  
LOCUS AC099052 139280 bp DNA linear PRI 27-APR-2002  
DEFINITION Homo sapiens chromosome 3 clone RP11-437M11, complete sequence.  
ACCESSION AC099052 AC013495  
VERSION AC099052.2 GI:20334583  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 139280)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
Direct Submission  
TITLE

3334	3332	8696	8706	3211	3274	4501	4494	1539	1576	801	810
2067	2130	6	<800	6382	6509	1246	1196	9653	9683	2871	2870
7056	7041	1909	1855	512	<800	3506	3541	1139	1106	6483	6509
192	<800	9283	9141	449	<800	26	<800	4988	5046	475	<800
6704	6557	2412	2548	4573	4461	8	<800	6671	6733	13148	12835
1550	1550	1706	1657	3848	3827	2587	2707	10647	10644	2634	2667
592	<800	13289	13045	2253	2291	1786	1875	3923	3899	5084	5150
4161	4208	4390	4360	2063	2080	9945	9829	148	<800	3053	3135
646	<800	5556	5570	10555	10560	2384	2406	3510	3495	692	<800
11112	10883	59	<800	5319	5359	566	<800	5156	5046	11	<800
2333	2406	5316	5307	1692	1676	13468	14157	3082	3095	1217	1197
1312	1311	2688	2730	1757	1676	1736	1875	2190	2180	18955	19193
1020	1040	202	<800	2791	2870	1667	1699	3488	3495		
9291	9095	1403	1383	1656	1676	4212	4208	16611	16955		
4538	4494	5062	5046	7317	7406	429	<800	834	849		
9281	9095	7618	7632	3977	4047	15	<800	1592	1657		
2275	2406	3644	3628	356	<800	75	<800	1072	1053		
4461	4494	1051	1053	10777	10560	1058	1040	11533	11404		
621	<800	6005	6000	8263	8317	158	<800	1417	1504		
1199	1196	102	<800	2997	3026	557	<800				
4463	4494	6161	6223	1027	1025						
9798	9829	3848	3899	9772	9719						
8361	8316	180	<800	460	<800						
327	<800	776	761	3648	3668						
1032	1040	1245	1197	16661	16928						
1918	1875	552	<800	5968	6078						
1003	1040	2099	2180	150	<800						
7003	7041	3224	3219	6146	6078						
3271	3332	871	849	2639	2667						
331	<800	2537	2548	13119	12835						
376	<800	745	<800	502	<800						
443	<800	4622	4657	110	<800						
1790	1875	2156	2180	2363	2371						
2604	2707	1211	1197	9088	8966						
912	931	1686	1657	1847	1777						
238	<800	9850	9683	1581	1525						

Query Match 93.6%; Score 23.4; DB 8; Length 139280;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
|||||  
Db 133475 AAAAAAAAAATCTCGGCCAGCATGG 133451

RESULT 7  
AC099405 142943 bp DNA linear PRI 11-JAN-2003  
LOCUS  
DEFINITION Papio anubis clone RP41-161K2, complete sequence.  
AC099405  
AC099405.3 GI:27597007  
VERSION  
KEYWORDS HTG.  
SOURCE Papio anubis (olive baboon)  
ORGANISM Papio anubis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Cercopithecoidea; Cercopithecinae; Papio.  
1 (bases 1 to 142943)  
Akhmer, N., Anconellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Carlagia, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,  
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,  
Laric, P., Lee-Lin, S.-O., Legaspi, R., Maduro, O.L., Maduro, V.B.,  
Marquies, E.H., Masello, C., Maskeri, B., McDowell, J.,  
Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,  
Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,  
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,  
Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Green,E.D.
Direct Submission
Submitted (14-NOV-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 142943)
Green,E.D.
Direct Submission
Submitted (26-FEB-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
4 (bases 1 to 142943)
Green,E.D.
Direct Submission
Submitted (11-JAN-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jan 11, 2003 this sequence version replaced gi:18921308.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@ngri.nih.gov
----- Project Information
Center project name: csk
Center clone name: 161K02

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the features section.
FEATURES
source
1. 142943
/organism="Papio anubis"
/mol_type="genomic DNA"
/db_xref="taxon:9555"
/clone="RP41-161K2"
/clone_lib="RP41"
<1..2206
/note="this sequence is not the entire insert of clone
RP41-161K2; clone overlaps with GenBank Accession Number
AC098807 (nucleotides 186894-212643) clone RP41-157N15
(center project name csk); this annotated segment
represents overlap with nucleotides 210438-212643 of
AC098807"
misc_feature
142488..142943
/note="clone has a very small overlap with GenBank
Accession Number AC099744 (nucleotides 1-456) clone
RP41-65E12 (center project name csk)"

ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 142943;
Best Local Similarity 96.0%; Pred. No. 2.5;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
|||||
Db 62967 AAAAAAAAAATCAGCGCCAGGCATGG 62991
|||||

RESULT 8
AC004593 150221 bp DNA linear PRI 30-JAN-2004
LOCUS AC004593 Homo sapiens PAC clone RP5-964C11 from 7, complete sequence.

```

```

ACCESSION AC004593
VERSION AC004593.1
KEYWORDS GI:3063518
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 150221)
AUTHORS Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
Wyllie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,M.E., Cordes,M.,
Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
Dauphin-Kohlberg,S., Kozlowski-Reilly,A., Leonard,S., Rohlfing,T.,
Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
Strommatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
Woesner,J.P., Wendt,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Mohlmann,P.E.,
Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadonato,S.,
Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
The DNA sequence of human chromosome 7
Nature 424 (6945), 157-164 (2003)
12853948
REFERENCE 2 (bases 1 to 150221)
AUTHORS Becker,M., Hinds,K. and Yoakum,M.
TITLE The sequence of Homo sapiens PAC clone RP5-964C11
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 150221)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 150221)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 150221)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2004) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_DU0964C11
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

## SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTA-325M7, 200 bp overlap.  
Actual start of this clone is at base position 1 of RP5-964C11.  
Actual end is at 150025 of RP5-964C11.

## FEATURES

## Location/Qualifiers

1..150221  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="7"  
/map="7"

/clone="RP5-964C11"

/clone\_lib="RPCI-5"

51..106

/rpt\_family="(TAGA)n"

128..263

/rpt\_family="(TA)n"

281..729

/rpt\_family="ERV1"

1480..1656

/rpt\_family="MER1\_type"

2211..2467

/rpt\_family="Alu"

2487..2555

/rpt\_family="GA-rich"

2851..2982

/rpt\_family="MIR"

3234..3272

/rpt\_family="(TG)n"

3349..3662

/rpt\_family="Alu"

3790..3921

/rpt\_family="MIR"

4105..4275

/rpt\_family="MIR"

4431..4456

/rpt\_family="AT-rich"

4856..5158

/rpt\_family="Alu"

5179..5248

/rpt\_family="L2"

6844..7138

/rpt\_family="Alu"

7211..7356

/rpt\_family="L2"

8537..8726

/rpt\_family="MIR"

9996..11337

/notes="CpG island (%GC=71.5, o/e=0.79, #CpGs=139)"

10151..10180

/rpt\_family="GC-rich"

10615..10897

gene

## mRNA

/gene="CHN2"

10615..10897

/gene="CHN2"

/rpt\_family="(CGG)n"

10755..10857

10849..>10897

/gene="CHN2"

/note="Homo sapiens chimera (Chimaerin) 2 (CHN2), mRNA.;

H DJ0964C11.1

This gene was based on gi(4757979)

Continues as H RG325M07.2"

/codon\_start=1

/product="unknown"

/protein\_id="AA02045.1"

/db\_xref="GI:41393527"

/translation="MAASSNSLSGSSVSS"

11503..11576

/rpt\_family="(GA)n"

13112..13164

/rpt\_family="MIR"

13240..13436

/rpt\_family="MIR"

13568..13905

/rpt\_family="L2"

16200..16377

/rpt\_family="CT-rich"

17089..17221

/rpt\_family="MIR"

17492..17681

/rpt\_family="MER1\_type"

18897..19091

/rpt\_family="MER1\_type"

19092..19399

/rpt\_family="Alu"

19400..19528

/rpt\_family="MER1\_type"

19975..20023

/rpt\_family="(TG)n"

21304..21353

/rpt\_family="MIR"

21400..21447

/rpt\_family="L2"

21805..21980

/rpt\_family="L1"

21987..22077

/rpt\_family="L2"

22221..22486

/rpt\_family="Alu"

22683..23040

/rpt\_family="L1"

24937..25140

/rpt\_family="MeLR"

## Query Match

Best Local Similarity 93.6%; Score 23.4; DB 8; Length 150221;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCAGGCATGG 25

Db 111375 AAAAAAAAAATCAGCCAGGCATGG 111399

## RESULT 9

AC144870

LOCUS

DEFINITION

AC144870

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC144870 163473 bp DNA linear HTG 09-JUN-2003  
Pan troglodytes clone CH251-288H14, WORKING DRAFT SEQUENCE, 7  
ordered pieces.

AC144870.2 GI:31544081  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan.  
 1 (bases 1 to 163473)  
 Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K.,  
 Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,  
 Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P.,  
 Hurlb, B., Idol, J.R., Karling, E., Kwong, P., Laric, P., Lee-Lin, S.-O.,  
 Legaapi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masiello, C.,  
 Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E.,  
 Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,  
 Sison, C., Stantripop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,  
 Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 163473)  
 Green, E.D.  
 Direct Submission  
 Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 163473)  
 Green, E.D.  
 Direct Submission  
 Submitted (09-JUN-2003) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA  
 On Jun 9, 2003 this sequence version replaced gi:31044283.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: nisc.zoc@nhgri.nih.gov  
 ----- Project Information  
 Center project name: esd  
 Center clone name: 288H14

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 162199 bases at least Q40  
 Consensus quality: 162591 bases at least Q30  
 Consensus quality: 162777 bases at least Q20  
 Insert size: 149000; agarose-fp  
 Insert size: 162873; sum-of-contigs  
 Quality coverage: 10.65x in Q20 bases; agarose-fp  
 Quality coverage: 9.74x in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 1 49711: contig of 49711 bp in length  
 49712 49811: gap of unknown length  
 49812 93246: contig of 43435 bp in length  
 93247 93347: gap of unknown length  
 93347 126704: contig of 33358 bp in length

\* 126705 126804: gap of unknown length  
 \* 126805 128476: contig of 1672 bp in length  
 \* 128477 128576: gap of unknown length  
 \* 128577 143981: contig of 15405 bp in length  
 \* 143982 144081: gap of unknown length  
 \* 144082 145826: contig of 1745 bp in length  
 \* 145827 145926: gap of unknown length  
 \* 145927 163473: contig of 17547 bp in length.  
 FEATURES  
 source  
 1..163473  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="CH251-288H14"  
 /clone\_lib="CH251"  
 1..60595  
 /notes="clone overlaps with GenBank Accession Number AC096875 clone RP43-120N4 (center project name cl2)"  
 1..49711  
 /notes="assembly\_fragment  
 clone\_end:T7  
 vector\_side:left"  
 49712..49811  
 /estimated\_length=unknown  
 49812..93246  
 /notes="assembly\_fragment"  
 93247..93346  
 /estimated\_length=unknown  
 93347..126704  
 /notes="assembly\_fragment"  
 103034..163473  
 /notes="clone overlaps with GenBank Accession Number AC097009 clone RP43-169E8 (center project name cma)"  
 126705..126804  
 /estimated\_length=unknown  
 126805..128476  
 /notes="assembly\_fragment"  
 128477..128576  
 /estimated\_length=unknown  
 128577..143981  
 /notes="assembly\_fragment"  
 143982..144081  
 /estimated\_length=unknown  
 144082..145826  
 /notes="assembly\_fragment"  
 145827..145926  
 /estimated\_length=unknown  
 145927..163473  
 /notes="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:right"  
 ORIGIN  
 Query Match 93.6%; Score 23.4; DB 14; Length 163473;  
 Best Local Similarity 96.0%; Pred. No. 2.5;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAATCAGGCGCAGGCATGG 25  
 |||||  
 Db 45133 AAAAAAAAAATCAGGCGCAGGCATGG 45157  
 RESULT 10  
 AL135914/c  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP1-56L9 on chromosome 6 Contains a novel gene, the SOD2 gene for superoxide dismutase 2, mitochondrial, a pseudogene similar to heterogeneous nuclear ribonucleoprotein H, the gene for Wilms' tumor 1-associating protein (KIAA0105) (WTAP), the ACAT2 gene for acetyl-Coenzyme A acetyltransferase 2 (acetoacetyl Coenzyme A thiolase), the TCPI gene for t-complex 1 (CTC1, Cctt, D6S230E), the 5' end of the MRPL18 gene for mitochondrial ribosomal protein L18, and 4 CpG





```

/evidence=not experimental
/rpt family="AluJb"
18465..16486
/evidence=not experimental
/rpt family="AT rich"
16492..16803
/evidence=not experimental
/rpt family="AluY"
16838..17049
/evidence=not experimental
/rpt family="AluJo"
17247..17332
/evidence=not experimental
/rpt family="MLT1K"
17851..17998
/evidence=not experimental
/rpt family="L2"
complement(18114..18296)
/evidence=not experimental
/rpt family="MIR"
complement(18503..18635)
/evidence=not experimental
/rpt family="LIM4"
complement(18851..18934)
/evidence=not experimental
/rpt family="LIP4"
19285..19356
/evidence=not experimental
/rpt family="L2"
19391..19697
/evidence=not experimental
/rpt family="AluSc"
19708..19987
/evidence=not experimental
/rpt family="AluSg"
19988..20022
/evidence=not experimental
/rpt family="(TAA)n"
20035..20363
/evidence=not experimental
/rpt family="AluJo"
20558..20711
/evidence=not experimental
/rpt family="MLT1H"
20807..21027
/evidence=not experimental
/rpt family="AluJb"
21028..21062
/evidence=not experimental
/rpt family="AluY"
complement(21074..21375)
/evidence=not experimental
/rpt family="AluJo"
complement(21472..21611)
/evidence=not experimental
/rpt family="AluSg/x"
complement(21612..21623)
/evidence=not experimental
/rpt family="AluY"
21624..21643
/evidence=not experimental
/rpt family="(TTTA)n"
complement(21644..21913)
/evidence=not experimental
/rpt family="AluY"
21914..22185
/evidence=not experimental
/rpt family="AluJb"
22190..22505
/evidence=not experimental
/rpt family="AluSq"
22510..22812
/evidence=not experimental

```

```

repeat_region      /rpt family="AluJb"
22872..23141
/evidence=not experimental
/rpt family="AluSg1"
complement(23261..23563)
/evidence=not experimental
/rpt family="AluJo"
23741..24077
/evidence=not experimental
/rpt family="AluJb"
24095..24405
/evidence=not experimental
/rpt family="AluY"
24719..24774
/evidence=not experimental
/rpt family="(CATATA)n"
24915..25015
/evidence=not experimental
/rpt family="AT rich"
complement(25077..25132)
/evidence=not experimental
/rpt family="MLT1D"
25191..25245
/evidence=not experimental
/rpt family="MER66A"
complement(25556..25720)
/evidence=not experimental
/rpt family="LIM4"
25717..26058
/evidence=not experimental
/rpt family="LIME1"
26150..26446
/evidence=not experimental
/rpt family="AluSx"
27486..27783
/evidence=not experimental
/rpt family="AluSx"
27982..28268

```

Query Match 93.6%; Score 23.4; DB 8; Length 168813;  
 Best Local Similarity 96.0%; Pred. No. 2.5;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCATGG 25  
 |||||

Db 38340 AAAAAAAAAATCAGCGCATGG 38316

# RESULT 12

AC146225/c

LOCUS AC146225 171272 bp DNA linear PRI 18-DEC-2003

DEFINITION Pan troglodytes BAC clone RP43-32E8 from 7, complete sequence.

ACCESSION AC146225

VERSION AC146225.2 GI:39725897

KEYWORDS HTG.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Pan.

REFERENCE 1 (bases 1 to 171272)

AUTHORS Van Brunt, A., Bielicki, L. and Haglund, K.

TITLE The sequence of Pan troglodytes BAC clone RP43-32E8

JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 171272)

AUTHORS Wilson, R.

TITLE Sequencing of Pan troglodytes

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 171272)

AUTHORS Wilson, R.K.

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444  
 Forest Park Parkway, St. Louis, MO 63108, USA

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
4 (bases 1 to 171272)
Wilson,R.K.
Direct Submission
Submitted (11-DEC-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 171272)
Wilson,R.
Direct Submission
Submitted (18-DEC-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 11, 2003 this sequence version replaced gi:3338178.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
-----
Center project name: C_PT032E08
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, 'Cint', Verkes #C0471; birthdate: 6-6-80). The
clone and detailed information can be obtained from Resgen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1..171272
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clone="RP43-32E8"
/clone_lib="RPCI-43"
98712..99115
misc_feature
/note="Sequence derived from PCR product of project DNA."

ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 171272;
Best Local Similarity 96.0%; Pred. NO. 2.5;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCCGCATGG 25
DB 5317 AAAAAAAAAATCAGCCCGCATGG 5293

RESULT 13
AC092041

```

```

LOCUS
DEFINITION
AC092041 Homo sapiens chromosome 3 clone RP11-132K4, complete sequence.
AC092041 AC018471
VERSION
AC092041.2 GI:16554366
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
1 (bases 1 to 171461)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 171461)
Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
Haugen,E.D.
Direct Submission
Submitted (16-JUN-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 171461)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
Haugen,E.D.
Direct Submission
Submitted (31-OCT-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Oct 31, 2001 this sequence version replaced gi:14475974.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: BCM
----- Project Information
Center project name: Chr-3
Center clone name: RP11-132K4 (bc0251)
----- Summary Statistics
Sequencing vector: M13; L08821; 46% of reads
Sequencing vector: plasmid; L08752; 54% of reads
Chemistry: Dye-primer Bodipy; 21% of reads
Chemistry: Dye-terminator Eri; 41% of reads
Chemistry: Dye-terminator Big Dye; 37% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 170617 bases at least Q40
Consensus quality: 171425 bases at least Q30
Consensus quality: 171461 bases at least Q20
Insert size: 172100; 9.1% error; agarose-fp
Insert size: 169288; sum-of-contigs
Quality coverage: 6.4x in Q20 bases; agarose-fp
Quality coverage: 6.5x in Q20 bases; sum-of-contigs
-----
Overlapping Sequences:
5': RP11-296G21 (UWGC:bc0347) AC022064
3': Mapping in progress
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

```

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BcoRI				HindIII				BglII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8718	6678	6413	4913	4847						
6	<800	6382	6413	2067	2044						
2156	2188	512	<800	6965	7043						
1211	1199	449	<800	331	<800						
1686	1660	3307	3321	376	<800						
9850	10016	5968	5988	443	<800						
1539	1512	150	<800	1790	1807						
9653	10016	6146	6413	2604	2723						
1139	1096	2639	2658	912	932						
4989	5165	13120	13188	238	<800						
6671	7108	502	<800	4501	4488						
10647	10016	110	<800	1246	1203						
3923	3833	2363	2357	3506	3476						
148	<800	9088	9105	26	<800						
3510	3471	1847	1776	8	<800						
5156	5165	1581	1535	2587	2618						
3082	3060	801	822	1786	1807						
2190	2188	2871	2888	9946	10100						
3488	3471	6483	6413	2384	2295						
16610	16826	475	<800	566	<800						
834	851	13148	13188	13468	13603						
1592	1584	2634	2658	1736	1807						
1072	1096	5083	5080	1667	1625						
11533	11673	3053	3133	4212	4175						
1414	1403	692	<800	429	<800						
13506	13186	11	<800	15	<800						
13438	13186	1217	1225	75	<800						
7021	7108	18955	19298	1058	1074						

4626	4653	7378	7543	158	<800
373	<800	2204	2219	557	<800
660	<800	3558	3549	1941	1945
2077	2188	10428	10124	1851	1807
192	<800	6333	6413	1358	1325
565	<800	2867	2888	1100	1074
2741	2713	7646	7543	10141	10100
2032	2053	2371	2357	2785	2974
14931	15070	1276	1225	930	932
5203	5165	1104	1106	1836	1807
		15297	14858	2679	2832
		3433	3549	2950	2974
				3238	3265
				4244	4175
				106	<800
				5422	5449
				3226	3265
				3810	3851
				2893	2974
				1623	1625
				14167	13603
				8682	8622
				1609	1625
				12735	13603
				12350	11745
				2422	2478
				5492	5449

Query Match 93.6%; Score 23.4; DB 8; Length 171461;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCAGGCATGG 25  
|||||||  
Db 136541 AAAAAAAAAATCAGCGCAGGCATGG 136565

## RESULT 14

AC010653  
LOCUS AC010653 184515 bp DNA linear PRI 06-AUG-2002  
DEFINITION Homo sapiens chromosome 16 clone RP11-417N10, complete sequence.  
ACCESSION AC010653  
VERSION AC010653.10 GI:22122867  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 184515)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 184515)  
DOE Joint Genome Institute.  
Direct Submission

Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 184515)  
DOE Joint Genome Institute.  
Direct Submission

Submitted (06-AUG-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Aug 6, 2002 this sequence version replaced gi:17737037.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

-----

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

-----

FEATURES

source

1. 184515

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="RP11-417N10"

-----

ORIGIN

Query Match 93.6%; Score 23.4; DB 8; Length 184515;

Best Local Similarity .96.0%; Pred. No. 2.5;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCCGAGGATGG 25

|||||

Db 176048 AAAAAAAAAATCAGCCGAGGATGG 176072

|||||

AC006059

AC006059

AC006059.3 GI:4544348

HTG.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 184515)

Murphy, D., Aronson, A.D., Bouck, J., Bunac, C., Chen, J., Chen, Z.,

Culpepper, P., Ding, Y., Dugan, S.P., Durbin, K.J., Forcum, J.,

Ganesh, R.P., Garcia, C., Garcia, D.K., Gorrell, H., Gorrell, L.L.,

-----

REFERENCE

AUTHORS

185161

185162

0.000104921

0.0374023

20

0

-----

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base

quality are listed below. Description of the metrics can be found

at URL:

http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

He, X., Hernandez, J., Jackson, L.E., Kondejewski, N., Leal, B.,  
Lichtarge, O., Liu, W., Logan, O., Lu, J., Martinez, C., Moore, S.,  
Moorish, T., Nguyen, N., Oswal, G., Pampall, L.R., Parish, B.J.,  
Perez, L.M., Rashid, N.D., Rives, C.M., Scherer, S.E., Shen, H.,  
Simon, M.L., Vo, Q.K., Wei, Y., Williamson, A.L., Worley, K., Zhou, X.,  
Naylor, S.L., and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 185161)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 185161)  
Worley, K.C.  
Direct Submission

Submitted (30-MAR-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

4 (bases 1 to 185161)  
Worley, K.C.  
Direct Submission

Submitted (01-APR-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Mar 30, 1999 this sequence version replaced gi:4263627.

INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email

gc-help@bcm.tmc.edu

-----

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

-----

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST

(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the

EST and cDNA sequences. Genes demonstrate at least two exons

flanked by consensus splice sites that maintained sequence

continuity across the splice junctions. Sequences that are not

identical matches are annotated as similar.

-----

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum

standard of double strand coverage with a minimum of 2 clones and 2

reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for

a region does not meet this standard, it will be indicated in the

annotation as Low Coverage.

-----

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base

quality are listed below. Description of the metrics can be found

at URL:

http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

-----

QUALSTAT-REPORT-----

-----

Summary Statistics -----

Contig length: 185161

Phrap values in estimate: 185162

Average error rate (BCM-Phrap estimate): 0.000104921

Fraction of Phrap values less than 40: 0.0374023

Number of consensus changing edits: 20

Number of N's in consensus: 0





TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

NISC Comparative Sequencing Initiative  
 2 (bases 1 to 187530)  
 Green, E.D.  
 Direct Submission  
 Submitted (29-SEP-2001) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 187530)  
 Green, E.D.  
 Direct Submission  
 Submitted (22-MAY-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 On May 22, 2002 this sequence version replaced gi:15808565.  
 -----  
 Project Information  
 -----  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc.zoo@nhgri.nih.gov](mailto:nisc.zoo@nhgri.nih.gov)  
 -----  
 Project Information  
 -----  
 Center project name: clz  
 Center clone name: 120N04

The sequence data in this record represents an 'enhanced'  
 version of a Phase 2 submission. Specifically, the indicated  
 order and orientation of each sequence contig has been  
 established using one or more of the following: read-pair  
 data from individual subclones, overlaps with neighboring  
 clones, alignment with available reference sequence (e.g.,  
 human), and/or confirmation by PCR testing. In addition,  
 the sequence assembly is based on at least 8X average  
 coverage in Q20 bases and has been reviewed to rule out  
 gross misassemblies, the low-quality ends of sequence  
 contigs have been trimmed away, and each base is associated  
 with a Phrap-derived quality score.

#### ----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 18653 bases at least Q40  
 Consensus quality: 186859 bases at least Q30  
 Consensus quality: 187025 bases at least Q20  
 Insert size: 154000; agarose-fp  
 Insert size: 187130; sum-of-contigs  
 Quality coverage: 11.97x in Q20 bases; agarose-fp  
 Quality coverage: 9.85x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submittor.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 1 5374: contig of 5374 bp in length  
 \* 5375 5474: gap of unknown length  
 \* 5475 57679: contig of 5205 bp in length  
 \* 57680 57779: gap of unknown length  
 \* 57780 99879: contig of 42100 bp in length  
 \* 99880 99979: gap of unknown length  
 \* 99980 121286: contig of 21307 bp in length  
 \* 121287 121386: gap of unknown length  
 \* 121387 187530: contig of 66144 bp in length.

#### FEATURES

source  
 1. 187530  
 /organism="Pan troglodytes"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="RP43-120N4"  
 /clone.lib="RP43"  
 misc\_feature  
 1. 5374  
 /note="assembly\_fragment"

clone\_end:T7  
 vector\_side:left  
 5375..5474  
 /estimated\_length=unknown  
 5475..57679  
 /note="assembly\_fragment"  
 57680..57779  
 /estimated\_length=unknown  
 57780..99879  
 /note="assembly\_fragment"  
 99880..99979  
 /estimated\_length=unknown  
 99980..121286  
 /note="assembly\_fragment"  
 121287..121386  
 /estimated\_length=unknown  
 121387..187530  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right

#### ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 187530;  
 Best Local Similarity 96.8%; Pred. NO. 2.5;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
 |||||  
 Db 171950 AAAAAAAAAATCAGCGCCAGGCATGG 171974

#### RESULT 17

AC009097/c 199471.bp DNA linear PRI 19-MAR-2003  
 LOCUS Homo sapiens chromosome 16 clone RP11-43215, complete sequence.  
 DEFINITION AC009097  
 ACCESSION AC009097  
 VERSION AC009097.9 GI:29124044  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE 1 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
 Alamos National Laboratory.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission  
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission  
 JOURNAL Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission  
 JOURNAL Submitted (01-MAY-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 6 (bases 1 to 199471)  
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
 Alamos National Laboratory.

TITLE Direct Submission  
 JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA  
On Mar 19, 2003 this sequence version replaced gi:20376957.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
Location/Qualifiers  
1. .199471  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-43215"

ORIGIN  
Query Match 93.6%; Score 23.4; DB 8; Length 199471;  
Best Local Similarity 96.0%; Pred. No. 2.5;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATGG 25  
|||||||  
Db 147272 AAAAAAAAAATCAGCGCCAGCGCATGG 147248  
|||||||

RESULT 18  
AL590137 328439 bp DNA linear HTG 29-JUN-2002  
LOCUS Homo sapiens chromosome 1 clone RP11-23017, WORKING DRAFT SEQUENCE,  
DEFINITION 62 unordered pieces.  
ACCESSION AL590137 AC026065  
VERSION AL590137.7 GI:21665915  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 328439)  
McLay, K.  
Direct Submission  
Submitted (28-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 2, 2002 this sequence version replaced gi:21530915.  
Draft Sequence Produced by Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
http://genome.wustl.edu/gsc/index.shtml  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA23017  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M7815; 66% of reads  
Sequencing vector: plasmid; L08752; 33% of reads  
Chemistry: Dye-terminator; 5% of reads  
Chemistry: Dye-primer-amersham; 64% of reads  
Chemistry: Dye-terminator ET-amersham; 0% of reads  
Dye-terminator Big Dye; 29% of reads  
Consensus quality: 295728 bases at least Q40  
Consensus quality: 310124 bases at least Q30  
Consensus quality: 316292 bases at least Q20  
Insert size: 322339; sum-of-contigs  
Insert size: 163622; 9.0% error; agarose-fp  
Quality coverage: 3.54x in Q20 bases; sum-of-contigs Quality  
coverage: 7.39x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 62 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 11330: contig of 11330 bp in length  
\* 11331: gap of 100 bp  
\* 11430: contig of 2086 bp in length  
\* 11431: gap of 100 bp  
\* 13516: contig of 3132 bp in length  
\* 13517: gap of 100 bp  
\* 16748: contig of 5599 bp in length  
\* 16749: gap of 100 bp  
\* 16848: contig of 100 bp  
\* 22447: gap of 100 bp  
\* 22448: contig of 2856 bp in length  
\* 22547: gap of 100 bp  
\* 22548: contig of 100 bp  
\* 25403: contig of 2169 bp in length  
\* 25404: gap of 100 bp  
\* 25503: contig of 2169 bp in length  
\* 25504: gap of 100 bp  
\* 27672: contig of 3451 bp in length  
\* 27673: gap of 100 bp  
\* 27773: contig of 3451 bp in length  
\* 31223: contig of 100 bp  
\* 31224: gap of 100 bp  
\* 31324: contig of 8498 bp in length  
\* 39821: gap of 100 bp  
\* 39822: contig of 2163 bp in length  
\* 39822: gap of 100 bp  
\* 42084: contig of 9980 bp in length  
\* 42085: gap of 100 bp  
\* 42185: contig of 3676 bp in length  
\* 52164: gap of 100 bp  
\* 52264: contig of 2107 bp in length  
\* 55940: gap of 100 bp  
\* 55941: contig of 3633 bp in length  
\* 58147: gap of 100 bp  
\* 58247: contig of 2782 bp in length  
\* 61880: gap of 100 bp  
\* 61881: contig of 2574 bp in length  
\* 64762: gap of 100 bp  
\* 64762: contig of 2574 bp in length  
\* 64862: gap of 100 bp  
\* 64863: contig of 2574 bp in length  
\* 67437: gap of 100 bp  
\* 67437: contig of 2793 bp in length  
\* 70329: gap of 100 bp  
\* 70330: contig of 3836 bp in length  
\* 70429: gap of 100 bp  
\* 74265: gap of 100 bp  
\* 74266: contig of 3894 bp in length  
\* 74365: gap of 100 bp  
\* 78259: contig of 10838 bp in length  
\* 78359: gap of 100 bp  
\* 81977: contig of 2876 bp in length  
\* 89197: gap of 100 bp  
\* 89198: contig of 2876 bp in length  
\* 92173: gap of 100 bp  
\* 92174: contig of 2667 bp in length  
\* 92274: gap of 100 bp  
\* 94941: contig of 4643 bp in length  
\* 95040: gap of 100 bp  
\* 95041: contig of 3847 bp in length  
\* 99683: gap of 100 bp  
\* 99784: contig of 5212 bp in length  
\* 103631: gap of 100 bp  
\* 103730: contig of 5212 bp in length  
\* 108942: gap of 100 bp  
\* 108943: contig of 2987 bp in length  
\* 112029: gap of 100 bp  
\* 112030: contig of 4013 bp in length  
\* 112130: gap of 100 bp  
\* 116142: contig of 7590 bp in length  
\* 116242: gap of 100 bp  
\* 123832: contig of 100 bp  
\* 123833: gap of 100 bp  
\* 123933: contig of 3677 bp in length  
\* 127609: gap of 100 bp  
\* 127709: contig of 7502 bp in length  
\* 127710: gap of 100 bp  
\* 135211: contig of 2318 bp in length  
\* 135212: gap of 100 bp  
\* 135311: contig of 2318 bp in length  
\* 137629: gap of 100 bp  
\* 137630: contig of 7562 bp in length  
\* 145291: gap of 100 bp  
\* 145292: contig of 3641 bp in length  
\* 145392: gap of 100 bp  
\* 149032: contig of 3641 bp in length  
\* 149033: gap of 100 bp

```
* 149133 155039: contig of 5907 bp in length
* 155040 155139: gap of 100 bp
* 158523 158523: contig of 3384 bp in length
* 158524 158623: gap of 100 bp
* 158624 158623: gap of 100 bp
* 161127 161127: contig of 2504 bp in length
* 161128 161227: gap of 100 bp
* 161228 165679: contig of 4452 bp in length
* 165680 165779: gap of 100 bp
* 165780 171609: contig of 5830 bp in length
* 171610 171709: gap of 100 bp
* 171710 174717: contig of 3008 bp in length
* 174718 174817: gap of 100 bp
* 174818 202623: contig of 27806 bp in length
* 202624 202723: gap of 100 bp
* 202724 207580: contig of 4857 bp in length
* 207581 207680: gap of 100 bp
* 207681 219602: contig of 11922 bp in length
* 219603 219703: gap of 100 bp
* 219703 222053: contig of 2351 bp in length
* 222054 222153: gap of 100 bp
* 222154 227017: contig of 4864 bp in length
* 227018 227117: gap of 100 bp
* 227118 232205: contig of 5088 bp in length
* 232206 232305: gap of 100 bp
* 232306 235573: contig of 3268 bp in length
* 235574 235673: gap of 100 bp
* 235674 237804: contig of 2131 bp in length
* 237805 237904: gap of 100 bp
* 237905 240134: contig of 2230 bp in length
* 240135 242023: contig of 11789 bp in length
* 242024 252123: gap of 100 bp
* 252124 252124: contig of 7368 bp in length
* 252125 259492: gap of 100 bp
* 259493 259592: contig of 2418 bp in length
* 259593 262109: gap of 100 bp
* 262110 264611: contig of 2502 bp in length
* 264612 264711: gap of 100 bp
* 264712 289527: contig of 4816 bp in length
* 289528 289628: gap of 100 bp
* 289629 275447: contig of 5820 bp in length
* 275448 275547: gap of 100 bp
* 275548 282359: contig of 6812 bp in length
* 282360 282459: gap of 100 bp
* 282460 286378: contig of 3919 bp in length
* 286379 286478: gap of 100 bp
* 286479 291251: contig of 4773 bp in length
* 291252 291351: gap of 100 bp
* 291352 298415: contig of 7065 bp in length
* 298417 298516: gap of 100 bp
* 298517 305839: contig of 7323 bp in length
* 305840 305939: gap of 100 bp
* 305940 309254: contig of 3315 bp in length
* 309255 309354: gap of 100 bp
* 309355 312006: contig of 2652 bp in length
* 312007 312106: gap of 100 bp
* 312107 325800: contig of 13694 bp in length
* 325801 325900: gap of 100 bp
* 325901 328439: contig of 2539 bp in length.
FEATURES
    source
        1..328439
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="RP11-23017"
            /clone_lib="RPC1-11.1"
            1..11330
                /note="assembly fragment:01560
                fragment_chain:1"
                11431..13516
                    /note="assembly fragment:02337
                    fragment_chain:1"
misc_feature
    /note="assembly fragment:00007
    fragment_chain:1"
    16849..22447
        /note="assembly fragment:03052
        fragment_chain:2"
        22548..25403
            /note="assembly fragment:00504
            fragment_chain:2"
            25504..27672
                Query Match 93.6%; Score 23.4; DB 14; Length 328439;
                Best Local Similarity 96.0%; Pred. No. 2.5;
                Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
                QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
                Db 13205 AAAAAAAAAATCAGGCCAGGCATGG 13229
RESULT 19
AB030316/c
LOCUS AB030316 12593 bp DNA linear ROD 19-OCT-2004
DEFINITION Mus musculus Pig-n gene for phosphatidylinositolglycan class N, complete cds.
ACCESSION AB030316 AB030309 AB030310 AB030311 AB030312 AB030313
VERSION AB030314 AB030315
KEYWORDS AB030316.2 GI:60391425
ORGANISM Mus musculus (house mouse)
SOURCE Mus musculus
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Hong,Y., Maeda,Y., Watanabe,R., Ohishi,K., Mishkind,M., Riezman,H. and Kinoshita,T.
TITLE pig-n, a mammalian homologue of yeast Mcd4p, is involved in transferring phosphoethanolamine to the first mannose of the glycosylphosphatidylinositol
J. Biol. Chem. 274 (49), 35099-35106 (1999)
JOURNAL 10574991
PUBMED 2 (bases 1 to 12593)
REFERENCE Hong,Y.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Yeongjin Hong, Research Institute for Microbial Diseases, Department of Immunoregulation; 3-1, Yamada-Oka, Suita, Osaka 565-0871, Japan
JOURNAL (E-mail:yjhong@biken.osaka-u.ac.jp, Tel:81-6-879-8329, Fax:81-6-875-5233)
COMMENT On or before Mar 1, 2005 this sequence version replaced gi:5668561, gi:5668562, gi:5668563, gi:5668564, gi:5668565, gi:5668566, gi:5668567, gi:5668568, gi:5668569.
FEATURES
    source
        1..12593
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /tissue_type="liver"
            /tissue_lib="mouse 129 liver genomic library"
            237..11714
                /genes="Pig-n"
                join(237..399,1212..1285,1572..1691,2728..2976,3792..3913,4547..4645,5355..5461,6225..6349,7499..7629,8168..8284,8668..8708,9049..9108,10944..11036,11655..11714)
                /genes="Pig-n"
                237..399
                    /genes="Pig-n"
                    /product="phosphatidylinositolglycan class N"
                    /number=1
                    945..1044
                        /estimated length=unknown
                        1212..1285
```

```
misc_feature 13617..16748
    /note="assembly fragment:00007
    fragment_chain:1"
    16849..22447
        /note="assembly fragment:03052
        fragment_chain:2"
        22548..25403
            /note="assembly fragment:00504
            fragment_chain:2"
            25504..27672
                Query Match 93.6%; Score 23.4; DB 14; Length 328439;
                Best Local Similarity 96.0%; Pred. No. 2.5;
                Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
                QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
                Db 13205 AAAAAAAAAATCAGGCCAGGCATGG 13229
RESULT 19
AB030316/c
LOCUS AB030316 12593 bp DNA linear ROD 19-OCT-2004
DEFINITION Mus musculus Pig-n gene for phosphatidylinositolglycan class N, complete cds.
ACCESSION AB030316 AB030309 AB030310 AB030311 AB030312 AB030313
VERSION AB030314 AB030315
KEYWORDS AB030316.2 GI:60391425
ORGANISM Mus musculus (house mouse)
SOURCE Mus musculus
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Hong,Y., Maeda,Y., Watanabe,R., Ohishi,K., Mishkind,M., Riezman,H. and Kinoshita,T.
TITLE pig-n, a mammalian homologue of yeast Mcd4p, is involved in transferring phosphoethanolamine to the first mannose of the glycosylphosphatidylinositol
J. Biol. Chem. 274 (49), 35099-35106 (1999)
JOURNAL 10574991
PUBMED 2 (bases 1 to 12593)
REFERENCE Hong,Y.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Yeongjin Hong, Research Institute for Microbial Diseases, Department of Immunoregulation; 3-1, Yamada-Oka, Suita, Osaka 565-0871, Japan
JOURNAL (E-mail:yjhong@biken.osaka-u.ac.jp, Tel:81-6-879-8329, Fax:81-6-875-5233)
COMMENT On or before Mar 1, 2005 this sequence version replaced gi:5668561, gi:5668562, gi:5668563, gi:5668564, gi:5668565, gi:5668566, gi:5668567, gi:5668568, gi:5668569.
FEATURES
    source
        1..12593
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /tissue_type="liver"
            /tissue_lib="mouse 129 liver genomic library"
            237..11714
                /genes="Pig-n"
                join(237..399,1212..1285,1572..1691,2728..2976,3792..3913,4547..4645,5355..5461,6225..6349,7499..7629,8168..8284,8668..8708,9049..9108,10944..11036,11655..11714)
                /genes="Pig-n"
                237..399
                    /genes="Pig-n"
                    /product="phosphatidylinositolglycan class N"
                    /number=1
                    945..1044
                        /estimated length=unknown
                        1212..1285
```

exon	/gene="Pig-n" /product="phosphatidylinositolglycan class N" /number=2		/gene="Pig-n" /number=14	
	1572. .1691		Query Match 89.6%; Score 22.4; DB 9; Length 12593; Best Local Similarity 92.0%; Pred. No. 7.6; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
exon	/gene="Pig-n" /number=3			
	2728. .2976			
CDS	/gene="Pig-n" /number=4			
	Join(2756. .2976,3792. .3913,4547. .4645,5355. .5461, 6225. .6349,7499. .7629,8168. .8284,8668. .8708,9049. .9108, 10944. .11036,11655. .11714)		1 AAAAAAAAAATCAGCCAGGCATGG 25       3290 AAAAAAAAAATRANGCCAGGCATGG 3266 	
	/gene="Pig-n" /note="homology to yeast Saccharomyces cerevisiae MCD4"			
	/codon_start=1 /product="phosphatidylinositolglycan class N" /protein_id="BAA82663.1" /db_xref="GI:5668571"		RESULT 20 BV092806/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	
exon	/translation="MLLPALGLLIHFVFFASIFDIYPTSLVHGMTPQFTPLPPPAK RLVLFVADGLRADTLIELDEDCNSRAPFIRNVIMHGSWGVSHTVPTESRPGHVALI AGFYEDVSAAVGKXENPVEFDSLNFESKITWSGSPDILPMFAGAGSDHVIITYSD AQREFGADATKJDTWDFDKVDFDFAARNQSIFTKVNEEKVVFLLHLGIDTNGH AHRPSRREYKIKKVPQNSAQOYDDEFLEKRWLENKRDVNOADIAPLMASLIGV STLTPFTVWGAIKFPQNSAQOYDDEFLEKRWLENKRDVNOADIAPLMASLIGV PPLASVGILPVGLNNTGLFKAESMFTNAVQLLEQFKVQTKKEATLPLFLTTFPK" 3792. .3913		Mus musculus Mus musculus (house mouse) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus. 1 (bases 1 to 394) Usuka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J. D., Foernzler, D. and Peltz, G. Mus musculus SNPs Unpublished (2003)	
exon	/gene="Pig-n" /number=5			
	4547. .4645		Contact: Jonathan Usuka Roche Palo Alto Genetics and Genomics Department Roche Palo Alto 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA Tel: 6508555807 Email: Jonathan.Usuka@roche.com Primer A: No primer submitted. Primer B: No primer submitted. Location/Qualifiers	
gap	/number=6			
	5081. .5180		1. .394 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /map="1-13610-13406-CAA01163581.1.1.19206" /clone_lib="Roche Palo Alto" /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, B10.D2-H2/OsNj, BALB/cByJ, BALB/cJ, 129/Sv, AKR/J, B10.D2-H2/OsNj, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei. "	
exon	/estimated_length=unknown /gene="Pig-n" /number=7			
	5927. .6026		FEATURES source	
exon	/estimated_length=unknown /gene="Pig-n" /number=8			
	6225. .6349		Query Match 87.2%; Score 21.8; DB 10; Length 394; Best Local Similarity 92.0%; Pred. No. 15; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
gap	/estimated_length=unknown /gene="Pig-n" /number=9			
	7873. .7972		QY 1 AAAAAAAAAATCAGCCAGGCATGG 25       163 AAAAAAAAAATCTTGGCCAGGCATGG 139 	
exon	/estimated_length=unknown /gene="Pig-n" /number=10			
	8373. .8472		STS ORIGIN	
gap	/estimated_length=unknown /gene="Pig-n" /number=11			
	8668. .8708			
exon	/estimated_length=unknown /gene="Pig-n" /number=12			
	9049. .9108			
gap	/estimated_length=unknown /gene="Pig-n" /number=13			
	9781. .9880			
exon	/estimated_length=unknown /gene="Pig-n" /number=14			
	10944. .11036			
exon	/estimated_length=unknown /gene="Pig-n" /number=15			
	11655. .11710			

```

VERSION BV099496.1 GI:40812576
KEYWORDS STS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
1 Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 398)
USUKA,J., LIAO,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
TITLE Mus musculus SNPs
JOURNAL Unpublished (2003)
COMMENT
Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.
Location/Qualifiers
FEATURES
source
1..398
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="1-13610-13406-CAAA01163581.1.1.19206"
/clone_lib="Roche Palo Alto"
/notes="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/cSnJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei."
<1..398
STS
ORIGIN
Query Match 87.2%; Score 21.8; DB 10; Length 398;
Best Local Similarity 92.0%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
|||||
Db 168 AAAAAAAAAATCTTGGCCAGGCATGG 144
|||||
RESULT 23
AF270524/c
LOCUS Homo sapiens clone 2002c9t3_SMARCA5/SHGC-33076 map 13qtel sequence.
DEFINITION AF270524
ACCESSION AF270524
VERSION AF270524.2 GI:11863208
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 701)
REFERENCE Riethman,H.C. and Moyzis,R.K.
AUTHORS Integration of telomeric DNA sequences with the human reference
TITLE sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 701)
AUTHORS Riethman,H.C. and Moyzis,R.K.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2000) Molecular Genetics, The Wistar Institute,
3601 Spruce St., Philadelphia, PA 19104, USA
REFERENCE 3 (bases 1 to 701)
AUTHORS Riethman,H.C. and Moyzis,R.K.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2000) Molecular Genetics, The Wistar Institute,
3601 Spruce St., Philadelphia, PA 19104, USA
REMARK Sequence update by submitter
COMMENT On Dec 15, 2000 this sequence version replaced gi:9755066.
FEATURES
source
1..701
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13qtel"
/clone="2002c9t3_SMARCA5/SHGC-33076"
ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 701;
Best Local Similarity 92.0%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
|||||
Db 60 AAAAAAAAAATCTTGGCCAGGCATGG 36
|||||
RESULT 24
BV69872/c

```





/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..5861

# STS ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 861;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
|||||  
Db 788 AAAAAAAAAATCAGCGCCAGCATGG 764

RESULT 26  
BV594288  
LOCUS  
DEFINITION  
SV594288  
tagged site.  
BV594288  
SV594288.1 GI:62510017  
STS.  
SOURCE  
ORGANISM  
Pan troglodytes verus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.  
1 (bases 1 to 942)  
Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
Jaffe,D.B.  
Initial Sequence of the Chimpanzee Genome and Comparison with the  
Human Genome  
Unpublished (2005)

CONTACT: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 942

Protocol:  
23,021,928 chimpanzee whole genome shotgun reads were aligned to  
the Human genome NCBI  
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
including Clint (Pan  
troglodytes verus), 3 other Pan troglodytes verus chimps  
(Donald, Karlén, Yvonne), 3 Pan  
troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps  
of unknown origin  
(Gon Unknown Chimp). Common names: Pan troglodytes verus is the  
western chimp and Pan  
troglodytes troglodytes is the central chimp. To be included in  
chimpanzee SNP discovery, a  
read must be at least 500bp in length, at least 50% of its base  
calls must have Phred  
score >= 20, at least 30% of its base calls must satisfy  
SNQS(30,25) (single strand NQS, the  
base in question has Phred score >= 30, the surrounding 10 bases in  
the read have Phred  
score >= 25), and the read must have at least 200 bp SNQS(30,25)  
bases. Reads not uniquely  
placed in the genome and read pairs whose two ends were not  
consistently placed were  
discarded. After above filtering, NQS(30,25) standard was applied  
to all pairs of  
overlapping reads to call NQS bases and SNPs. Alignments (between  
two reads) with less  
than 100 NQS bases or with SNP rate > 0.01 were discarded. To

exclude alignment between two  
copies of a single read, comparisons between two reads that share  
95% of their genome  
alignments (>=95% bases of read A and >=95% bases of read B were  
placed at the same locus  
of human genome) were discarded.  
Location/Qualifiers  
1..942  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..5942

# FEATURES

source  
1..942  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..5942

# STS ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 942;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
|||||  
Db 106 AAAAAAAAAATCAGCGCCAGCATGG 130

# RESULT 27

AK090515/c  
LOCUS  
DEFINITION  
Homo sapiens cDNA FLJ33196 fis, clone ADRGL2006034.  
AK090515  
ACCESSION  
AK090515.1 GI:21748691  
VERSION  
oligo capping; fis (full insert sequence).  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1  
REFERENCE  
AUTHORS  
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,  
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,  
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,  
Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T.,  
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
Tashiro, H., Tanigami, A., Fujiiwara, T., Ono, T., Yamada, K., Fujii, Y.,  
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,  
Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,  
Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,  
Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.

Complete sequencing and characterization of 21,243 full-length  
human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
14702039  
2  
REFERENCE  
AUTHORS  
Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,

Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
3 (bases 1 to 2141)  
Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI); (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES  
Source

Location/Qualifiers  
1..2141  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ADRL2006034"  
/tissue\_type="adrenal gland"  
/clone\_lib="ADRL2"  
/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 2141;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATGG 25

DB 1169 AAAAAATAATATCAGCGCCAGCGCATGG 1145

RESULT 28  
AK097083/c

LOCUS Homo sapiens cDNA FLJ39764 fis, clone SPLEN2000143. PRI 30-JAN-2004

ACCESSION AK097083

VERSION AK097083.1 GI:21756734

oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwavanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, S., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Inose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Teraashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kunagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiya, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikeda, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.  
Complete sequencing and characterization of 21,243 full-length human cDNAs

TITLE  
JOURNAL

PUBMED

REFERENCE

AUTHORS

Nat. Genet. 36 (1), 40-45 (2004)

14702039

2

Unpublished

3 (bases 1 to 2530)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI); (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

Location/Qualifiers

1..2530

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="SPLEN2000143"

/tissue\_type="spleen"

/clone\_lib="SPLEN2"

/note="cloning vector: pME18SFL3"

FEATURES

source

1..2530

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="SPLEN2000143"

/tissue\_type="spleen"

/clone\_lib="SPLEN2"

/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 2530;

Best Local Similarity 92.0%; Pred. No. 15;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATGG 25

DB 1135 AAAAAATAATATCAGCGCCAGCGCATGG 1111

RESULT 29

AK097092/c

LOCUS Homo sapiens cDNA FLJ39764 fis, clone SPLEN2000143. PRI 30-JAN-2004

ACCESSION AK097092

VERSION AK097092.1 GI:1478335

oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3041)

REFERENCE

AUTHORS

3041 bp mRNA linear ROD 03-AUG-1996

acyl-coenzyme A:cholesterol acyltransferase [mice, peritoneal macrophages, mRNA, 3041 nt].

S81092

GI:1478335

Accession

Version

Keywords

Source

Organism

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3041)

REFERENCE

AUTHORS

3041 bp mRNA linear ROD 03-AUG-1996

acyl-coenzyme A:cholesterol acyltransferase [mice, peritoneal macrophages, mRNA, 3041 nt].

S81092

GI:1478335

Accession

Version

Keywords

Source

Organism

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3041)

Green, S., Steinberg, D. and Quehenberger, O.  
Cloning and expression in xenopus oocytes of a mouse homologue of  
the human acyl-coenzyme A: cholesterol acyltransferase and its  
potential role in metabolism of oxidized LDL  
Biochem. Biophys. Res. Commun. 218 (3), 924-929 (1996)  
GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 175976] from the original journal article.

## FEATURES

source  
1..3041  
/organism="Mus sp."  
/mol\_type="mRNA"  
/db\_xref="taxon:10095"  
1..3041  
/gene="acyl-coenzyme A:cholesterol acyltransferase, ACAT"  
146..1768  
/genes="acyl-coenzyme A:cholesterol acyltransferase, ACAT"  
/note="ACAT"  
/codon\_start=1  
/product="acyl-coenzyme A:cholesterol acyltransferase"  
/protein\_id="AAB36050.1"  
/db\_xref="GI:11478336"  
/translation="MSLRNRLSKSGENPQDEAQKNPMDTYRNGHITMKQIAKGRLL  
AAAEELKPLFMKEVGCHDFDFTNLIKESASLDGGCALTFPSILEMKONHRAKDL  
RAPPEQKIFISQSLDELFEVDHIRTIVHMFIALILFVLSITVVDYIDEGRLVLE  
FNLLAYAFGKFTVIMTWAMFSLTSLIPYLFQWAGHYSKSHPLIYSLVHGLLPL  
VQGLGVLGFTVIVLAYTLPPASRPILLLEQIRLIMKAHSFVRENI PRLVNAKES  
SKDPLTNVQYLYFELFAPTLIYRDNTPTPTVRMGVAMQFLQVFCGLFVYIYFERL  
CAPLFRNIKQEPFSARVLVLCVNSILPGVLILFLSFFAPLHCLWNAFAEMLRFGDRM  
FYKDWNNSTYSNYRTVNVVHDMLYVYVYKDLLWFFSKRFKSAAMLAVFALSAYVH  
EVALAICLSFYFVPLFVLFPMFGMAFNFIWDSRKRPIWNIMVWASLFLGYGLICFY  
SQSWARQCPLKNPFLDYVRPTWTCTRVF"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 3041;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCCGAGCATGG 25  
|||||  
Db 2776 AAAAAAAAAATCTTGGCCAGCATGG 2752  
|||||

RESULT 30  
MUSACACT  
LOCUS  
DEFINITION  
MUS musculus acyl-coenzyme A:cholesterol acyltransferase (ACACT)  
mRNA, complete cds.  
L42293  
L42293.1 GI:1066809  
acyl-coenzyme A:cholesterol acyltransferase.  
MUS musculus (house mouse)  
MUS musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 3697)  
Uelmen, P.J., Oka, K., Sullivan, M., Chang, C.C., Chang, T.Y. and  
Chan, L.  
Tissue-specific expression and cholesterol regulation of  
acyl-coenzyme A:cholesterol acyltransferase (ACAT) in mice.  
Molecular cloning of mouse ACAT cDNA, chromosomal localization, and  
regulation of ACAT in vivo and in vitro  
J. Biol. Chem. 270 (44), 26192-26201 (1995)  
7592824  
Original source text: Mus musculus cDNA to mRNA.  
Location/Qualifiers  
1..3697  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/map="chromosome 1"  
/note="(vector lambda ZAP11)"

## FEATURES

## source

Gene  
1..3697  
/gene="ACACT"  
mRNA  
1..3697  
/gene="ACACT"  
5'UTR  
1..809  
/gene="ACACT"  
TATA\_signal  
753..760  
/gene="ACACT"  
exon  
761..>1016  
/gene="ACACT"  
number=1  
810..2432  
/gene="ACACT"  
/note="amino acid feature: N-glycosylation site, aa 399;  
amino acid feature: N-glycosylation site, aa 481"  
/codon\_start=1  
/product="acyl-coenzyme A: cholesterol acyltransferase"  
/protein\_id="AAC42075.1"  
/db\_xref="GI:1066810"  
/translation="MSLRNRLSKSGENPQDEAQKNPMDTYRNGHITMKQIAKGRLL  
AAAEELKPLFMKEVGCHDFDFTNLIKESASLDGGCALTFPSILEMKONHRAKDL  
RAPPEQKIFISQSLDELFEVDHIRTIVHMFIALILFVLSITVVDYIDEGRLVLE  
FNLLAYAFGKFTVIMTWAMFSLTSLIPYLFQWAGHYSKSHPLIYSLVHGLLPL  
VQGLGVLGFTVIVLAYTLPPASRPILLLEQIRLIMKAHSFVRENI PRLVNAKES  
SKDPLTNVQYLYFELFAPTLIYRDNTPTPTVRMGVAMQFLQVFCGLFVYIYFERL  
CAPLFRNIKQEPFSARVLVLCVNSILPGVLILFLSFFAPLHCLWNAFAEMLRFGDRM  
FYKDWNNSTYSNYRTVNVVHDMLYVYVYKDLLWFFSKRFKSAAMLAVFALSAYVH  
EVALAICLSFYFVPLFVLFPMFGMAFNFIWDSRKRPIWNIMVWASLFLGYGLICFY  
SQSWARQCPLKNPFLDYVRPTWTCTRVF"  
2433..3696  
/gene="ACACT"  
3'UTR  
3697  
polyA\_site  
/gene="ACACT"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 3697;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCCGAGCATGG 25  
|||||  
Db 3444 AAAAAAAAAATCTTGGCCAGCATGG 3420  
|||||

RESULT 31  
BX571793/c  
LOCUS  
DEFINITION  
Human DNA sequence from clone W12-801302G5 on chromosome 1,  
complete sequence.  
BX571793  
BX571793.2 GI:33412544  
HTG.  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 35616)  
Barlow, K.

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----

This sequence was finished as follows unless otherwise noted: all chemistry were either double-stranded or sequenced with an alternate 30'; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 W12-801302G5 is from a Whitehead human fosmid library VECTOR: pEpifos-5 Eco721.

## FEATURES

source

## Location/Qualifiers

1..35616  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="W12-801302G5"  
 /clone\_lib="Whitehead\_fosmid"  
 2000  
 /note="Clone\_right\_end: RP11-5P18"

## misc\_feature

## ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 35616;  
 Best Local Similarity 92.0%; Pred. NO. 15;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 10247 AAAAAAAAAAGCAAGCGCCAGGCATGG 10223

## RESULT 32

## AC104230

LOCUS Homo sapiens chromosome 8 clone RP11-351N23 map 8, LOW-PASS  
 DEFINITION HTG 06-DEC-2001

AC104230 46784 bp DNA linear HTG 06-DEC-2001

## SEQUENCE SAMPLING.

AC104230

AC104230.1 GI:17386398

HTG; HTGS\_PHASE0.

KEYWORDS Homo sapiens

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 46784)

Birren,B., Linton,L., Nusbaum,C. and Lander,B.

Homo sapiens chromosome 8, clone RP11-351N23

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguski,M., Bouckgeer,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collumore,A., Cook,A., Cooke,P., DeArallano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mianga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,B., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L21645

Center clone name: 351\_N\_23

-----

\* NOTE: This record contains 57 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 699: contig of 699 bp in length

\* 700 799: gap of 100 bp

\* 800 1509: contig of 710 bp in length

\* 1510 1509: gap of 100 bp

\* 1610 2310: contig of 701 bp in length

\* 2311 2410: gap of 100 bp

\* 2411 3155: contig of 745 bp in length

\* 3156 3255: gap of 100 bp

\* 3256 3986: contig of 731 bp in length

\* 3987 4086: gap of 100 bp

\* 4087 4806: contig of 720 bp in length

\* 4807 4906: gap of 100 bp

\* 4907 5644: contig of 738 bp in length

\* 5645 5744: gap of 100 bp

\* 5745 6461: contig of 717 bp in length

\* 6462 6561: gap of 100 bp

\* 6562 7280: contig of 719 bp in length

\* 7281 7380: gap of 100 bp

\* 7381 8097: contig of 716 bp in length

\* 8097 8196: gap of 100 bp

\* 8197 8921: contig of 725 bp in length

\* 8922 9021: gap of 100 bp

\* 9022 9736: contig of 715 bp in length

\* 9737 9836: gap of 100 bp

\* 9837 10571: contig of 735 bp in length

\* 10572 10671: gap of 100 bp

\* 10672 11412: contig of 741 bp in length

\* 11413 11512: gap of 100 bp

\* 11513 12355: contig of 743 bp in length

\* 12356 12555: gap of 100 bp

\* 12556 13067: contig of 712 bp in length

\* 13068 13167: gap of 100 bp

\* 13168 13913: contig of 746 bp in length

\* 13914 14013: gap of 100 bp

\* 14014 14754: contig of 741 bp in length

\* 14755 14854: gap of 100 bp

\* 14855 15554: contig of 700 bp in length

\* 15555 15654: gap of 100 bp

\* 15655 16368: contig of 714 bp in length

\* 16369 16468: gap of 100 bp

\* 16469 17309: contig of 741 bp in length

\* 17310 17309: gap of 100 bp

\* 17310 18055: contig of 746 bp in length

\* 18056 18155: gap of 100 bp

TITLE  
 JOURNAL

COMMENT

/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-351N23"  
/clone\_lib="RPC1-11 Human Male BAC"  
700..799  
/estimated\_length=100  
1510..1609  
/estimated\_length=100  
2311..2410  
/estimated\_length=100  
3156..3255  
/estimated\_length=100  
3987..4086  
/estimated\_length=100  
4807..4906  
/estimated\_length=100  
5645..5744  
/estimated\_length=100

Query Match 87.2%; Score 21.8; DB 14; Length 46784;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
|||||||  
DB 33028 AAAAAAAAAATCAGCGCCAGCATGG 33052

RESULT 33  
AL358393  
LOCUS  
DEFINITION  
AL358393 50175 bp DNA linear PRI 18-MAY-2005  
Human DNA sequence from clone Rp11-540K16 on chromosome 1 Contains  
the 3' end of a novel gene (LOC284646) and a novel gene, complete  
sequence.  
AL358393  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 50175)  
Almeida, J.  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Nov 16, 2001 this sequence version replaced gi:15722103.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
Rp11-540K16 is from the library RPC1-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such

18156 18877: contig of 722 bp in length  
\* 18878 18977: gap of 100 bp  
\* 18979 19077: contig of 730 bp in length  
\* 19078 19807: gap of 100 bp  
\* 19808 20545: contig of 738 bp in length  
\* 20546 21351: contig of 100 bp  
\* 21352 21451: gap of 100 bp  
\* 21452 22171: contig of 720 bp in length  
\* 22172 22271: gap of 100 bp  
\* 22272 22989: contig of 718 bp in length  
\* 22990 23089: gap of 100 bp  
\* 23090 23801: contig of 712 bp in length  
\* 23802 23901: gap of 100 bp  
\* 23902 24637: contig of 736 bp in length  
\* 24638 24737: gap of 100 bp  
\* 24738 25495: contig of 758 bp in length  
\* 25496 25595: gap of 100 bp  
\* 25596 26286: contig of 691 bp in length  
\* 26287 26386: gap of 100 bp  
\* 26387 27116: contig of 730 bp in length  
\* 27117 27216: gap of 100 bp  
\* 27217 27919: contig of 703 bp in length  
\* 27920 28019: gap of 100 bp  
\* 28020 28767: contig of 748 bp in length  
\* 28768 28867: gap of 100 bp  
\* 28868 29589: contig of 722 bp in length  
\* 29590 30402: contig of 713 bp in length  
\* 30403 30502: gap of 100 bp  
\* 30503 31182: contig of 680 bp in length  
\* 31183 31282: gap of 100 bp  
\* 31283 31977: contig of 695 bp in length  
\* 31978 32077: gap of 100 bp  
\* 32078 32772: contig of 695 bp in length  
\* 32773 32872: gap of 100 bp  
\* 32874 33586: contig of 714 bp in length  
\* 33587 33686: gap of 100 bp  
\* 33687 34430: contig of 744 bp in length  
\* 34431 34530: gap of 100 bp  
\* 34531 35269: contig of 739 bp in length  
\* 35270 35369: gap of 100 bp  
\* 35370 36086: contig of 717 bp in length  
\* 36087 36186: gap of 100 bp  
\* 36187 36896: contig of 710 bp in length  
\* 36897 36996: gap of 100 bp  
\* 36997 37707: contig of 711 bp in length  
\* 37708 37807: gap of 100 bp  
\* 37808 38559: contig of 752 bp in length  
\* 38560 38659: gap of 100 bp  
\* 38660 39352: contig of 693 bp in length  
\* 39353 39452: gap of 100 bp  
\* 39453 40153: contig of 701 bp in length  
\* 40154 40253: gap of 100 bp  
\* 40254 40970: contig of 717 bp in length  
\* 40971 41070: gap of 100 bp  
\* 41071 41819: contig of 749 bp in length  
\* 41820 41919: gap of 100 bp  
\* 41920 42645: contig of 726 bp in length  
\* 42646 42745: gap of 100 bp  
\* 42746 43443: contig of 698 bp in length  
\* 43444 43543: gap of 100 bp  
\* 43544 44272: contig of 729 bp in length  
\* 44273 44372: gap of 100 bp  
\* 44373 45111: contig of 739 bp in length  
\* 45112 45211: gap of 100 bp  
\* 45212 45962: contig of 751 bp in length  
\* 45963 46062: gap of 100 bp  
\* 46063 46784: contig of 722 bp in length.

Location/Qualifiers  
1..46784  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

FEATURES  
Source



## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-357L18, 2000 bp overlap;  
the clone sequenced to the right is RP11-568024, 2000 bp overlap.  
Actual start of this clone is at base position 9509 of RP11-357L18;  
actual end is at base position 77842 of RP11-568024.

## FEATURES

```
Location/Qualifiers
1. .54618
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-350L15"
/clone_lib="RPCi-11"
30. .251
/rpt_family="L2"
repeat_region
369. .497
/rpt_family="MIR"
repeat_region
812. .1108
/rpt_family="Alu"
repeat_region
1332. .1767
/rpt_family="ERV1"
repeat_region
2574. .2655
/rpt_family="L1"
repeat_region
2656. .2685
/rpt_family="AT_rich"
repeat_region
2690. .3097
/rpt_family="L1"
repeat_region
3086. .3173
/rpt_family="L1"
repeat_region
3141. .3564
/rpt_family="L1"
repeat_region
3565. .5771
/rpt_family="L1"
repeat_region
5749. .5940
/rpt_family="L1"
repeat_region
6342. .6465
/rpt_family="MaLR"
repeat_region
7644. .7655
/rpt_family="L1"
repeat_region
7656. .8022
/rpt_family="L1"
repeat_region
8023. .8286
/rpt_family="L1"
repeat_region
8287. .8313
/rpt_family="(TTG)n"
repeat_region
8314. .8953
/rpt_family="L1"
repeat_region
9743. .9823
/rpt_family="MaLR"
repeat_region
10794. .10837
/rpt_family="AT_rich"
repeat_region
11545. .11873
/rpt_family="L1"
repeat_region
11900. .12152
/rpt_family="Alu"
repeat_region
12376. .12423
/rpt_family="(TATG)n"
repeat_region
12484. .12550
/rpt_family="L1"
repeat_region
12551. .12575
/rpt_family="(CAAAA)n"
repeat_region
12576. .13165
/rpt_family="L1"
repeat_region
13184. .13224
/rpt_family="L1"
repeat_region
13290. .13448
/rpt_family="Alu"
repeat_region
13506. .13531
/rpt_family="AT_rich"
repeat_region
15831. .16294
/rpt_family="L1"
repeat_region
16290. .16489
```

```
repeat_region /rpt_family="L1"
16568. .16735 /rpt_family="(TA)n"
repeat_region
16731. .16911 /rpt_family="(TTATA)n"
repeat_region
16925. .17098 /rpt_family="(TTATA)n"
repeat_region
17099. .17200 /rpt_family="(TA)n"
repeat_region
17204. .17654 /rpt_family="(TTATA)n"
repeat_region
17659. .18080 /rpt_family="(TATAA)n"
repeat_region
18086. .18245 /rpt_family="(TTATA)n"
repeat_region
18257. .18488 /rpt_family="(TTATA)n"
repeat_region
18493. .18558 /rpt_family="(TA)n"
repeat_region
18603. .18732 /rpt_family="(TA)n"
repeat_region
18765. .19018 /rpt_family="L1"
repeat_region
19548. .20379 /rpt_family="L1"
repeat_region
20463. .20486 /rpt_family="AT_rich"
repeat_region
20502. .20527 /rpt_family="AT_rich"
repeat_region
20642. .20820 /rpt_family="L1"
repeat_region
20950. .21613 /rpt_family="L1"
repeat_region
21614. .21923 /rpt_family="Alu"
repeat_region
21924. .22811 /rpt_family="L1"
repeat_region
22785. .23287 /rpt_family="L1"
repeat_region
23304. .23796 /rpt_family="L1"
repeat_region
23797. .24104 /rpt_family="Alu"
repeat_region
24105. .24151 /rpt_family="L1"
repeat_region
24145. .24914
```

Query Match 87.2%; Score 21.8; DB 8; Length 54618;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCAGCATGG 25  
|||||||  
Db 24117 AAAAAAAAAATCAGCGCAGCATGG 24093

## RESULT 35

AC105044  
LOCUS AC105044 Homo sapiens chromosome 8 clone CTD-3089F7 map 8, LOW-PASS SEQUENCE  
DEFINITION SAMPLING.  
ACCESSION AC105044.1 GI:17977423  
VERSION HTG; HTGS\_PHASE50.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
REFERENCE 1 (bases 1 to 54618)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 8, clone CTD-3089F7  
JOURNAL Unpublished



```

REFERENCE
AUTHORS
2 (bases 1 to 65882)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Katat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczkv,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23045
Center clone name: 3089_F_7
-----
* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 744: contig of 744 bp in length
* 745 844: gap of 100 bp
* 845 1563: contig of 719 bp in length
* 1564 1663: gap of 100 bp
* 1664 2395: contig of 732 bp in length
* 2396 2496: gap of 100 bp
* 2496 3230: contig of 735 bp in length
* 3231 3231: gap of 100 bp
* 3331 4071: contig of 741 bp in length
* 4072 4171: gap of 100 bp
* 4172 4911: contig of 740 bp in length
* 4912 5011: gap of 100 bp
* 5012 5745: contig of 734 bp in length
* 5746 5845: gap of 100 bp
* 5846 6556: contig of 711 bp in length
* 6557 6656: gap of 100 bp
* 6657 7362: contig of 706 bp in length
* 7363 7462: gap of 100 bp
* 7463 8196: contig of 734 bp in length
* 8197 8296: gap of 100 bp
* 8297 9005: contig of 709 bp in length
* 9006 9105: gap of 100 bp
* 9106 9846: contig of 741 bp in length
* 9847 9946: gap of 100 bp
*
* 9947 10642: contig of 696 bp in length
* 10643 10742: gap of 100 bp
* 10743 11488: contig of 746 bp in length
* 11489 11588: gap of 100 bp
* 11589 12322: contig of 734 bp in length
* 12323 12422: gap of 100 bp
* 12423 13136: contig of 714 bp in length
* 13137 13236: gap of 100 bp
* 13237 13975: contig of 739 bp in length
* 13976 14075: gap of 100 bp
* 14076 14803: contig of 728 bp in length
* 14804 14903: gap of 100 bp
* 14904 15608: contig of 705 bp in length
* 15609 15708: gap of 100 bp
* 15709 16452: contig of 744 bp in length
* 16453 16552: gap of 100 bp
* 16553 17300: contig of 748 bp in length
* 17301 17400: gap of 100 bp
* 17401 18118: contig of 718 bp in length
* 18119 18218: gap of 100 bp
* 18219 18937: contig of 719 bp in length
* 18938 19037: gap of 100 bp
* 19038 19776: contig of 739 bp in length
* 19777 19876: gap of 100 bp
* 19877 20609: contig of 733 bp in length
* 20610 20709: gap of 100 bp
* 20710 21422: contig of 713 bp in length
* 21423 21522: gap of 100 bp
* 21523 22208: contig of 686 bp in length
* 22209 23090: gap of 100 bp
* 23091 23190: contig of 782 bp in length
* 23191 23520: contig of 730 bp in length
* 23521 24021: contig of 100 bp
* 24021 24770: contig of 750 bp in length
* 24771 24870: gap of 100 bp
* 24871 25607: contig of 737 bp in length
* 25608 25707: gap of 100 bp
* 25708 26439: contig of 732 bp in length
* 26440 26539: gap of 100 bp
* 26540 27288: contig of 749 bp in length
* 27289 27388: gap of 100 bp
* 27389 28128: contig of 740 bp in length
* 28129 28228: gap of 100 bp
* 28229 28574: contig of 746 bp in length
* 28575 29074: gap of 100 bp
* 29075 29809: contig of 735 bp in length
* 29810 29909: gap of 100 bp
* 29910 30690: contig of 781 bp in length
* 30691 30790: gap of 100 bp
* 30791 31516: contig of 726 bp in length
* 31517 31616: gap of 100 bp
* 31617 32329: contig of 713 bp in length
* 32330 32429: gap of 100 bp
* 32430 33156: contig of 727 bp in length
* 33157 33256: gap of 100 bp
* 33257 34020: contig of 764 bp in length
* 34021 34120: gap of 100 bp
* 34121 34848: contig of 728 bp in length
* 34849 34948: gap of 100 bp
* 34949 35680: contig of 732 bp in length
* 35681 35780: gap of 100 bp
* 35781 36529: contig of 749 bp in length
* 36530 36629: gap of 100 bp
* 36630 37372: contig of 743 bp in length
* 37373 37472: gap of 100 bp
* 37473 38211: contig of 739 bp in length
* 38212 38311: gap of 100 bp
* 38312 39051: contig of 740 bp in length
* 39052 39151: gap of 100 bp
* 39152 39898: contig of 747 bp in length
* 39899 40735: contig of 737 bp in length
* 39999 40735: contig of 737 bp in length

```

40736 40835: gap of 100 bp  
\* 40836 41581: contig of 746 bp in length  
\* 41582 41681: gap of 100 bp  
\* 41682 42402: contig of 721 bp in length  
\* 42403 42502: gap of 100 bp  
\* 42503 43226: contig of 724 bp in length  
\* 43227 43326: gap of 100 bp  
\* 43327 44032: contig of 706 bp in length  
\* 44033 44132: gap of 100 bp  
\* 44133 44881: contig of 749 bp in length  
\* 44882 44981: gap of 100 bp  
\* 44982 45717: contig of 736 bp in length  
\* 45718 45817: gap of 100 bp  
\* 45818 46566: contig of 749 bp in length  
\* 46567 46666: gap of 100 bp  
\* 46667 47406: contig of 740 bp in length  
\* 47407 47506: gap of 100 bp  
\* 47507 48241: contig of 735 bp in length  
\* 48242 48341: gap of 100 bp  
\* 48342 49068: contig of 727 bp in length  
\* 49069 49168: gap of 100 bp  
\* 49169 49904: contig of 736 bp in length  
\* 49905 50004: gap of 100 bp  
\* 50005 50746: contig of 742 bp in length  
\* 50747 50846: gap of 100 bp  
\* 50847 51589: contig of 743 bp in length  
\* 51590 51689: gap of 100 bp  
\* 51690 52423: contig of 734 bp in length  
\* 52424 52523: gap of 100 bp  
\* 52524 53277: contig of 754 bp in length  
\* 53278 53377: gap of 100 bp  
\* 53378 54105: contig of 728 bp in length  
\* 54106 54205: gap of 100 bp  
\* 54206 54957: contig of 752 bp in length  
\* 54958 55057: gap of 100 bp  
\* 55058 55760: contig of 703 bp in length  
\* 55761 55860: gap of 100 bp  
\* 55861 56610: contig of 750 bp in length  
\* 56611 56710: gap of 100 bp

Query Match 87.2%; Score 21.8; DB 14; Length 65882;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
Db 9384 AAAAAAAAAATCAGCGCCAGCATGG 9408

RESULT 36  
AL589910/c  
LOCUS  
DEFINITION Human DNA sequence from clone Rp11-57L9 on chromosome 6 Contains part of the gene for a novel protein, complete sequence.  
ACCESSION AL589910  
VERSION AL589910.9 GI:14041767  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Garner, P.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
COMMENT On May 15, 2001 this sequence version replaced gi:13897415. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

40835: gap of 100 bp  
41581: contig of 746 bp in length  
41681: gap of 100 bp  
42402: contig of 721 bp in length  
42502: gap of 100 bp  
43226: contig of 724 bp in length  
43326: gap of 100 bp  
44032: contig of 706 bp in length  
44132: gap of 100 bp  
44881: contig of 749 bp in length  
44981: gap of 100 bp  
45717: contig of 736 bp in length  
45817: gap of 100 bp  
46566: contig of 749 bp in length  
46666: gap of 100 bp  
47406: contig of 740 bp in length  
47506: gap of 100 bp  
48241: contig of 735 bp in length  
48341: gap of 100 bp  
49068: contig of 727 bp in length  
49168: gap of 100 bp  
49904: contig of 736 bp in length  
50004: gap of 100 bp  
50746: contig of 742 bp in length  
50846: gap of 100 bp  
51589: contig of 743 bp in length  
51689: gap of 100 bp  
52423: contig of 734 bp in length  
52523: gap of 100 bp  
53277: contig of 754 bp in length  
53377: gap of 100 bp  
54105: contig of 728 bp in length  
54205: gap of 100 bp  
54957: contig of 752 bp in length  
55057: gap of 100 bp  
55760: contig of 703 bp in length  
55860: gap of 100 bp  
56610: contig of 750 bp in length  
56710: gap of 100 bp

Query Match 87.2%; Score 21.8; DB 8; Length 66591;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
Db 45932 AAAAAAAAAATTAAGCCAGCATGG 45908

RESULT 37  
AL133141  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone Rp11-421F9, complete sequence.

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6  
Rp11-57L9 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
Location/Qualifiers  
1. 66591  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="Rp11-57L9"  
/clone\_lib="RPI1-11.1"  
/notes="Clone left end: Rp11-57L9"  
join(complement(AL139098.15:434..575),  
complement(<43428..54443))  
/locus\_tag="Rp11-57L9.1-001"  
join(complement(AL139098.15:434..575),  
complement(<43428..54443),complement(50144..50321),  
complement(<43428..43496))  
/locus\_tag="Rp11-57L9.1-001"  
/notes="match: ESTg: AW502864 BF356516 BF356518 BF356520 BF356526 B149952"  
join(complement(AL139098.15:434..558),  
complement(54282..54443),complement(50144..50321),  
complement(<43428..43496))  
/locus\_tag="Rp11-57L9.1-001"  
/standard\_name="OTTHUMP0000017110"  
/codon\_start=1  
/protein\_id="CA116051.1"  
/db\_xref="GI:55960366"  
/db\_xref="UniProt/TREMBL:Q5SZM6"  
/translations="MLQAMLRLLFQSVKKEITGAPSLCAEAILLHLEETDENPHNVE  
FVKYLRQHIGNTLGSMLIEEMEKTSDRNQEGCYDVTVVQVTKRTOESKEYKEMMH  
YKKNIMIAVVEISMINKFEDETRNQERKKIQEKSHSYTDNCSDSDSSLSNQSKYKFC  
QKQLQILDLQDPGQPKR"  
63055  
/notes="Clone right end: Rp11-358H7"  
66492  
/note="Clone\_left\_end: RPI-276J11"

ORIGIN  
Query Match 87.2%; Score 21.8; DB 8; Length 66591;  
Best Local Similarity 92.0%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
Db 45932 AAAAAAAAAATTAAGCCAGCATGG 45908

RESULT 37  
AC133141  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone Rp11-421F9, complete sequence.





Query Match	87.2%;	Score 21.8;	DB 8;	Length 73084;
Best Local Similarity	92.0%;	Pred. No. 16;		
Matches	23;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY	1	AAAAAAAAATCAGGCCGAGGCATGG	25	
Db	3519	AAAAAAAAAACACGCCGAGGCATGG	3543	
RESULT 39				
AP000280				
LOCUS				
DEFINITION				
ACCESSION	AP000280			
VERSION	AP000280.3	GI:71891787		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.			
TITLE	Homo sapiens 73.686bp genomic DNA of 21q22.1			
JOURNAL	Published Only in Database (1999)			
REFERENCE	2 (bases 1 to 73687)			
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa, 228-8555, Japan [E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924]			
COMMENT	On Aug 5, 2005 this sequence version replaced gi:6016885. The sequence is a part of the data (ACCESSION No. AP000174 - AP000194). The sequencing project is supported by Japan Science Technology Corporation (JST) and The Institute of Physical and Chemical Research (RIKEN). Evidence from one overlapping clone (RP11-630H12), four mRNAs (BC069689, BC069709, BC069766, BC069779) and four ESTs (H11586, BX101811, AI888342, BX419841) indicate that there was an A missing from the chromosome 21 consensus sequence at position 29630 in this clone. We confirmed that this clone sequence was correct as previously submitted by reviewing the original trace data, but since all other evidence suggests that the base is truly missing we have inserted an additional base in this submission.			
FEATURES				
source	1. .73687			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	/chromosomes="21"			
	/map="21q22.1"			
	/clone="CMP21-S12"			
ORIGIN				
Query Match	87.2%;	Score 21.8;	DB 8;	Length 73687;
Best Local Similarity	92.0%;	Pred. No. 16;		
Matches	23;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY	1	AAAAAAAAATCAGGCCGAGGCATGG	25	
Db	35397	AAAAAAAAATCAGGCCGAGGCATGG	35421	
RESULT 40				
AC026374				
LOCUS				

complement (284469.1:44135. .44337),				
complement (284469.1:40468. .40797))				
/gene="WNK3"				
join(complement (6577. .7113),				
complement (AL049793.4:8687. .8859),				
complement (AL049793.4:6663. .6883),				
complement (AL049793.4:5490. .5647),				
complement (284469.1:143915. .144003),				
complement (284469.1:140319. .140538),				
complement (284469.1:136697. .136991),				
complement (284469.1:135278. .135471),				
complement (284469.1:135032. .135181),				
complement (284469.1:101418. .101482),				
complement (284469.1:97884. .98031),				
complement (284469.1:96103. .96219),				
complement (284469.1:93676. .93831),				
complement (284469.1:93543. .93594),				
complement (284469.1:92138. .92275),				
complement (284469.1:90841. .91243),				
complement (284469.1:80486. .80616),				
complement (284469.1:79082. .79695),				
complement (284469.1:74953. .75164),				
complement (284469.1:44135. .44337),				
complement (284469.1:40468. .40797))				
/gene="WNK3"				
/locus_tag="RP3-390O13.2-002"				
/product="WNK lysine deficient protein kinase 3"				
/note="match: cDNAs: AY082340.1"				
join(complement (6577. .7113),				
complement (AL049793.4:8687. .8859),				
complement (AL049793.4:6663. .6883),				
complement (AL049793.4:5490. .5647),				
complement (284469.1:143915. .144003),				
complement (284469.1:140319. .140538),				
complement (284469.1:136697. .136991),				
complement (284469.1:135278. .135471),				
complement (284469.1:135032. .135181),				
complement (284469.1:101418. .101482),				
complement (284469.1:97884. .98031),				
complement (284469.1:96103. .96219),				
complement (284469.1:93676. .93831),				
complement (284469.1:93543. .93594),				
complement (284469.1:92138. .92275),				
complement (284469.1:90841. .91243),				
complement (284469.1:80486. .80616),				
complement (284469.1:79082. .79695),				
complement (284469.1:74953. .75164),				
complement (284469.1:44135. .44337),				
complement (284469.1:40468. .40797))				
/gene="WNK3"				
/locus_tag="RP3-390O13.2-002"				
/standard_name="OTTHUMP0000023370"				
/note="match: proteins: Q8RCX6"				
/codon_start=1				
/product="WNK lysine deficient protein kinase 3"				
/protein_id="CAI40707.1"				
/db_xref="GI:57209549"				
/db_xref="GOA:Q5JRC2"				
/db_xref="InterPro:IPR000719"				
/db_xref="InterPro:IPR001245"				
/db_xref="InterPro:IPR002290"				
/db_xref="InterPro:IPR008271"				
/db_xref="UniProt/TrEMBL:Q5JRC2"				
/translation="MAIDSGDPASTDESKPDGISFNRPVQVAATLTVEARLKEKNS				
TFSSAGETVERKRFKFSKVSQEMTDDKVAESSPKDRIKAAINIPRVKLPFNSVLRGGQ				
EVKYEQSKSTSIKDCFKNEKEEMEEAAKAVATSPSGRFLKFDIELGRGAFKT				
VYKGLTEWVWACVLODRKLTAKAQORFKFEAEMLKGLQHPNIVRFYDWSIESILK				
GKKCIVLITELMTSGTLTKYTKRFPVKMPKVLKRSWCROILKGLQFLHTPTPIIHRDL				
KCDNIITITGPTGSKVIGDLGLATIMRTSFAKSVIGTPEFWAPMEYEHYDESVDVYAF				

DEFINITION	Homo sapiens 3 BAC RP11-336E10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION	AC026374
VERSION	AC026374.19 GI:18449530
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	<p>1 (bases 1 to 75385)</p> <p>Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrook,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Krtovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulberg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonri,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohitari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taylor,P., Tamerisa,K., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.</p>
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 75385)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 75385)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	<p>On Jan 31, 2002 this sequence version replaced gi:17223113.</p> <p>INFORMATION: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> or email <a href="mailto:gc-help@bcm.tmc.edu">gc-help@bcm.tmc.edu</a></p> <p>CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of</p>

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

	Location/Qualifiers
source	1..75385 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosomes="3" /clone="RP11-336E10"
misc_feature	1..2004 /note="overlaps bases 160026..162029 of clone AC020632"
repeat_region	2..553 /rpt_family="LIMB7"
repeat_region	579..877 /rpt_family="AluJb"
repeat_region	1065..1166 /rpt_family="LIP4A"
repeat_region	1167..1204 /rpt_family="AT rich" complement(1614..1665) /rpt_family="L2"
repeat_region	1625..1740 /rpt_family="MER5A" complement(2853..3084) /rpt_family="MIR"
repeat_region	8470..8502 /rpt_family="AT_rich"
repeat_region	8713..8844 /rpt_family="FLAM_A"
repeat_region	complement(9335..10455) /rpt_family="Tigger1"
repeat_region	complement(10465..10677) /rpt_family="LIP16"
repeat_region	complement(10678..11800) /rpt_family="Tigger1"
repeat_region	12385..12692 /rpt_family="AluSc"
repeat_region	12883..13188 /rpt_family="AluSp"
STS	13211..13348 /standard_name="31891"
repeat_region	complement(13989..14205)

```
repeat_region /rpt_family="MIR"
14885..14924
repeat_region /rpt_family="(CAAAA)n"
15402..15455
repeat_region /rpt_family="(TATG)n"
17217..17237
repeat_region /rpt_family="AT rich"
complement(17643..17709)
repeat_region /rpt_family="Alu"
complement(17831..18015)
repeat_region /rpt_family="AluSx"
18649..18830
repeat_region /rpt_family="MERS5C"
19196..19384
repeat_region /rpt_family="MERSA"
20219..20529
repeat_region /rpt_family="AluY"
complement(23001..23220)
repeat_region /rpt_family="MIR"
complement(23222..23307)
repeat_region /rpt_family="LiM4"
25299..25319
repeat_region /rpt_family="AT rich"
complement(25621..25798)
repeat_region /rpt_family="MIR"
complement(26067..26301)
repeat_region /rpt_family="MIR"
26322..26356
repeat_region /rpt_family="AT rich"
27319..27347
repeat_region /rpt_family="AT rich"
28341..28360
repeat_region /rpt_family="(CAAAA)n"
31655..31674
repeat_region /rpt_family="(A)n"
31679..31978
repeat_region /rpt_family="AluSg"
33125..33151
repeat_region /rpt_family="(A)n"
33795..34100
repeat_region /rpt_family="AluSx"
complement(34135..34447)
repeat_region /rpt_family="AluJo"
complement(36355..36645)
repeat_region /rpt_family="AluJo"
complement(37635..37928)
repeat_region /rpt_family="AluJb"
38025..38220
repeat_region /rpt_family="MIR"
39657..39694
repeat_region /rpt_family="AT rich"
complement(40476..40623)
repeat_region /rpt_family="AluJb"
```

```
Query Match 87.2%; Score 21.8; DB 8; Length 75385;
Best Local Similarity 92.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCAGGCATGG 25
|||||
Db 31666 AAAAAAAAAATTAGCGCAGGCATGG 31690
```

```
RESULT 41
HSDJ336M4/c 76241 bp DNA linear PRI 02-DEC-2004
LOCUS HSDJ336M4
DEFINITION Human DNA sequence from clone RP3-336M4 on chromosome 1p36.11-36.2
Contains the 5' UTR of the gene Miz-1, pseudogenes similar to part
of Carboxypeptidase D, fire oncogene and fire-2 oncogene (USP6), a
putative novel gene and the 3' end of the gene for a cardiovascular
heat shock protein. Contains ESTs, STSs, GSSs and two putative CpG
islands, complete sequence.
ACCESSION AL096775
```

```
AL096775.10 GI:7529574
HTG; CpG islands; heat shock protein; Miz-1; Tre; USP6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 76241)
Bird,C.
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 9, 2000 this sequence version replaced gi:6002162.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chri
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP3-336M4 is from the library RPCI-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-336M4. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP3-336M4 is at 76241 in this sequence.
The true right end of clone RP1-134019 is at 100 in this sequence.
FEATURES
Location/Qualifiers
1..76241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704M04336"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.11-36.2"
/clone="RP3-336M4"
/clone_lib="RPCI-3"
2..174
/note="AluYb8 repeat: matches 1..174 of consensus"
complement(183..359)
/note="match: STS: Em:G37055 Em:N75818"
609..935
/note="match: STS: Em:AL117303 Em:HS442J245"
629..912
/note="match: GSS: Em:AQ264259"
634..852
/note="match: GSS: Em:B74942"
634..833
```



```
misc_feature /note="match: GSS: Em:AQ311353"
634..768
/misc_feature /note="match: GSS: Em:AQ317636"
641..888
/misc_feature /note="match: GSS: Em:AQ678277"
complement(648..859)
/misc_feature /note="match: STS: Em:G17164"
648..813
/misc_feature /note="match: STS: Em:RSA108WE1"
complement(648..785)
/misc_feature /note="match: GSS: Em:AQ887510"
650..927
/misc_feature /note="match: GSS: Em:AQ869729"
651..859
/misc_feature /note="match: GSS: Em:AQ754783"
652..877
/misc_feature /note="match: GSS: Em:AQ282607"
complement(654..814)
/misc_feature /note="match: GSS: Em:AQ118318"
658..935
/misc_feature /note="match: GSS: Em:AQ376808"
661..912
/misc_feature /note="match: STS: Em:G40890"
complement(664..931)
/misc_feature /note="match: GSS: Em:AQ455403"
complement(676..877)
/misc_feature /note="match: GSS: Em:AQ933387"
679..898
/misc_feature /note="match: STS: Em:G01739"
680..884
/misc_feature /note="match: GSS: Em:AQ589116"
complement(680..877)
/misc_feature /note="match: GSS: Em:AQ203541"
684..889
/misc_feature /note="match: GSS: Em:AQ378981"
687..850
/misc_feature /note="match: GSS: Em:AQ538621"
688..893
/misc_feature /note="match: GSS: Em:AQ550571"
complement(688..819)
/misc_feature /note="match: GSS: Em:AQ337044"
690..908
/misc_feature /note="match: GSS: Em:AQ20968"
690..850
/misc_feature /note="match: GSS: Em:AQ341026"
695..833
/misc_feature /note="match: STS: Em:G33512 Em:RSA045WH9"
complement(708..934)
/misc_feature /note="match: GSS: Em:AQ777297"
complement(708..834)
/misc_feature /note="match: GSS: Em:AQ064770"
708..831
/misc_feature /note="match: GSS: Em:AQ387275"
complement(728..936)
/misc_feature /note="match: STS: Em:AQ268772"
complement(745..911)
/misc_feature /note="match: GSS: Em:B64703"
complement(746..864)
/misc_feature /note="match: GSS: Em:AQ148978"
776..938
/misc_feature /note="match: GSS: Em:AQ984702"
789..922
/misc_feature /note="match: GSS: Em:AQ198586"
789..911
/misc_feature /note="match: GSS: Em:AQ830058"
1025..1133
repeat_region /note="MIR repeat: matches 100..226 of consensus"
1496..1581
repeat_region /note="L2 repeat: matches 2613..2709 of consensus"
1850..2106
repeat_region /note="MLTJL repeat: matches 207..466 of consensus"
complement(join(2349..2435,5293..5409))
gene /gene="dJ336M4.1"

complement(join(<2349..2435,5293..5409))
/gene="dJ336M4.1"
/product="dJ336M4.1 (Miz-1)"
/note="match: CDNA: Em:U14556 Em:U22396 Em:Y09723
match: ESTs: Em:AA021389 Em:H31841 Em:AI655871 Em:AA624099
Em:AW478691 Em:AA018620 Em:AW408296 Em:AW294814
Em:AW432021 Em:AW481026"
/evidence=not_experimental
2539..2582
repeat_region /note="L2 repeat: matches 2648..2692 of consensus"
2583..2682
repeat_region /note="L2 repeat: matches 2576..2680 of consensus"
2842..3213
repeat_region /note="L2 repeat: matches 1731..2159 of consensus"
3218..3530
repeat_region /note="AluJo repeat: matches 1..312 of consensus"
3568..3783
repeat_region /note="AluJo repeat: matches 1..196 of consensus"
3793..3874
repeat_region /note="L2 repeat: matches 2181..2268 of consensus"
3883..4185
repeat_region /note="AluX repeat: matches 1..302 of consensus"
4490..4709
repeat_region /note="MIR repeat: matches 17..260 of consensus"
4718..5487
misc_feature /note="CpG island"
/evidence=not_experimental
5044..5255
repeat_region /note="106 copies 2 mer cc 58% conserved"
5576..5794
repeat_region /note="MIR repeat: matches 31..262 of consensus"
5925..6236
repeat_region /note="L2 repeat: matches 373..693 of consensus"
6321..6632
repeat_region /note="AluY repeat: matches 1..309 of consensus"
6646..6771
repeat_region /note="FLAM_A repeat: matches 1..133 of consensus"
6808..6930
repeat_region /note="LIPB2 repeat: matches 6032..6155 of consensus"
6931..7240
repeat_region /note="AluY repeat: matches 1..310 of consensus"
7241..7930
repeat_region /note="LIPB2 repeat: matches 5337..6032 of consensus"
7931..8240
repeat_region /note="AluX repeat: matches 1..301 of consensus"
8241..8966

Query Match 87.2%; Score 21.8; DB 8; Length 76241;
Best Local Similarity 92.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCAGCATGG 25
||||||| |||||
Db 50887 AAAAAAAAAAGCAAGCGCAGCATGG 50863

RESULT 42
AP0000039
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, segment 10/28,
DEFINITION complete sequence.
ACCESSION AP0000039
VERSION AP0000039.1 GI:3132349
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
TITLE Homo sapiens genomic DNA, chromosome 21q
JOURNAL Published Only in Database (1998)
```



chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries for genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-169113; the clone sequenced to the right is RP11-212J19, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-308G12; actual end is at base position 80717 of RP11-212J19.

Data from AC010823 and AC084013 was used to finish this clone, AC012356. Polymorphisms have been identified between AC010823, AC084013, and AC012356.

```

FEATURES
    source
        1. .112527
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="RP11-308G12"
            /clone_lib="RPCI-11"

ORIGIN
    Query Match      87.2%; Score 21.8; DB 8; Length 112527;
    Best Local Similarity 92.0%; Pred. No. 16;
    Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 69473 AAAAAAAAAATCAGCGCCAGGCATGG 69449

RESULT 47
HSJ635E18/c
LOCUS
DEFINITION
    Human DNA sequence from clone RP4-635E18 on chromosome
    1p36.11-36.31 Contains the TARDBP gene for TAR DNA binding protein,
    the MASP2 gene for mannan-binding lectin serine protease 2, the SRM
    gene for spermidine synthase, the PMSCL2 gene for 100kDa
    polyomitis/scleroderma autoantigen 2 (FLJ41371, FLJ36636, two
    novel genes (FLJ30609), the 3' end of the FRAP1 gene for FK506
    binding protein 12-rapamycin associated protein 1 and a CpG island,
    complete sequence.
AL109811
ACCESSION
    AL109811.40 GI:18152762
VERSION
    HTG; CpG island; FK506; FLJ36636; FLJ41371; FRAP1; MASP2; PMSCL2;
KEYWORDS
    SRM; TARDBP.
SOURCE
    Homo sapiens (human)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE
    1 (bases 1 to 112769)
    Wallis, J.
    Direct Submission

Sequence split into 4 fragments LOCUS AC145312 Accession AC145312
Fragment Name Begin End
AC145312.0 1 110000
AC145312.1 100001 210000
AC145312.2 200001 310000
AC145312.3 300001 408009
Continuation (2 of 4) of AC145312 from base 100001 (AC145312 Homo sapiens chromosome 16

Query Match      87.2%; Score 21.8; DB 14; Length 110000;
Best Local Similarity 92.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 13478 AAAAAAAAAATCAGCGCCAGGCATGG 13454

RESULT 46
AC012356/c
LOCUS
DEFINITION
    Homo sapiens BAC clone RP11-308G12 from 2, complete sequence.
ACCESSION
    AC012356
VERSION
    AC012356.6 GI:18093119
KEYWORDS
    HTG.
SOURCE
    Homo sapiens (human)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE
    1 (bases 1 to 112527)
    Haakenson, W. and Levy, A.
    The sequence of Homo sapiens BAC clone RP11-308G12
    Unpublished (2001)
REFERENCE
    2 (bases 1 to 112527)
    Waterston, R.H.
    Direct Submission
    Submitted (25-OCT-1999) Genome Sequencing Center, Washington
    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
    MO 63108, USA
REFERENCE
    3 (bases 1 to 112527)
    Waterston, R.
    Direct Submission
    Submitted (09-JAN-2002) Department of Genetics, Washington
    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
    4 (bases 1 to 112527)
    Wilson, R.K.
    Direct Submission
    Submitted (21-APR-2005) Genome Sequencing Center, Washington
    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
    MO 63108, USA
COMMENT
    On Jan 9, 2002 this sequence version replaced gi:16259110.
    -----
    Center: Washington University Genome Sequencing Center
    Center code: WUGSC
    Web site: http://genome.wustl.edu
    Contact: submissions@watson.wustl.edu
    -----
    Summary Statistics
    -----
    Center project name: H_NH0308G12
    -----

NOTICE:
    This sequence was finished as follows unless otherwise noted:
    all regions were double stranded, sequenced with an alternate
  
```

JOURNAL

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Jan 15, 2002 this sequence version replaced gi:11967852.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone configs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>  
 RP4-635E18 is from the library RP4-635E18 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

source

1. .112769  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosomes="1"  
 /map="p36.11-36.31"  
 /clone="RP4-635E18"  
 /clone\_lib="RPCI-4"

misc\_feature

1  
 /notes="Clone left end: RP4-635E18"

misc\_feature

2691. .3057  
 /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."  
 join(8705. .8776,10064. .10313,13192. .13355,15081. .15221,16777. .16837)  
 /genes="TARDBP"

gene

1  
 /locus tag="RP4-635E18.2-004"  
 join(8705. .8776,10064. .10313,13192. .13355,15081. .15221,16777. .16837)  
 /genes="TARDBP"

mRNA

1  
 /locus tag="RP4-635E18.2-004"  
 join(8705. .8776,10064. .10313,13192. .13355,15081. .15221,16777. .16837)  
 /genes="TARDBP"

/product="TAR DNA binding protein"

gene

1  
 /locus tag="RP4-635E18.2-001"  
 join(8990. .9091,10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .20391)  
 /genes="TARDBP"

mRNA

1  
 /locus tag="RP4-635E18.2-001"  
 join(8990. .9091,10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .20391)  
 /genes="TARDBP"

/product="TAR DNA binding protein"

gene

1  
 /locus tag="RP4-635E18.2-005"  
 join(9003. .9091,10064. .10313,11872. .11994,13192. .13355,15081. .15221)  
 /genes="TARDBP"  
 /locus tag="RP4-635E18.2-005"  
 join(9003. .9091,10064. .10313,11872. .11994,13192. .13355,15081. .15221)  
 /genes="TARDBP"

gene  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-005"  
 /product="TAR DNA binding protein"  
 /note="match: ESTs: Em:BM556331.1 Em:BX509845.1"  
 join(9020. .9091,10064. .10313,12471. .12602,13192. .13355,15081. .15221,16777. .16946)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-006"  
 join(9020. .9091,10064. .10313,12471. .12602,13192. .13355,15081. .15221,16777. .16946)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-006"  
 /product="TAR DNA binding protein"  
 /note="match: ESTs: Em:BE901725.1 Em:BG480879.1 Em:BG484362.1"  
 join(10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18598,19541. .19677,20412. .21367)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-002"  
 join(10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18598,19541. .19677,20412. .21367)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-002"  
 /product="TAR DNA binding protein"  
 /note="match: CDNAs: Em:AK045495.1 Em:AK076208.1"  
 join(10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18589,19541. .20531)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-003"  
 join(10064. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18589,19541. .20531)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-003"  
 /product="TAR DNA binding protein"  
 /note="match: ESTs: Em:BC021005.1"  
 join(10076. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18598,19541. .19596)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-002"  
 /standard\_name="OTHOMP00000002172"  
 /note="match: proteins: Tr:AAP92688 Tr:AAP92691 Tr:Q8R0B4"  
 /codon\_start=1  
 /product="TAR DNA binding protein"  
 /protein\_id="CAI22100.1"  
 /db\_xref="GI:56204105"  
 /db\_xref="GOA:Q5TEQ7"  
 /db\_xref="UniProt/TrEMBL:Q5TEQ7"  
 /translation="MSEYIRVTEDEDEPIEIPSEDDGTLLSTVTAQPPGACGLRYRNPVSCMGRLVEGILHAPDAGWNLVYVNYPKDKRKMDETASSAVKVRVQKTSDLVLGLPKTKTEQDLKEYFSTFGEVLMQVKDLKTHGSKGFGFVFTVEYQVKVNSQRHMDGRWCDCKLNSKQSDPELRSRVFVGRCTEDTDELREFFFSQYGDVMDFVIFPKPFAFAFVTFADDTQAQLCGEDLIKIGISVHISNAEPKHNSNRQLERSGRFGNPNVHLISNVYGRSTSLKVL"  
 join(10076. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18589,19541. .19596)  
 /gene="TARDBP"  
 /locus tag="RP4-635E18.2-003"  
 /standard\_name="OTHOMP00000002173"  
 /codon\_start=1  
 /product="TAR DNA binding protein"  
 /protein\_id="CAI22099.1"  
 /db\_xref="GI:56204104"  
 /db\_xref="GOA:Q5TEQ8"  
 /db\_xref="UniProt/TrEMBL:Q5TEQ8"  
 /translation="MSEYIRVTEDEDEPIEIPSEDDGTLLSTVTAQPPGACGLRYRNPVSCMGRLVEGILHAPDAGWNLVYVNYPKDKRKMDETASSAVKVRVQKTSDLVLGLPKTKTEQDLKEYFSTFGEVLMQVKDLKTHGSKGFGFVFTVEYQVKVNSQRHMDGRWCDCKLNSKQSDPELRSRVFVGRCTEDTDELREFFFSQYGDVMDFVIFPKPFAFAFVTFADDTQAQLCGEDLIKIGISVHISNAEPKHNSNRQLERSGRFGNPNVHLISNVYGRSTSLKVL"  
 join(10076. .10313,13192. .13355,15081. .15221,16777. .16947,18472. .18589,19541. .19596)  
 /gene="TARDBP"

CDS

```
18472..19002)
/locus tag="RP4-635E18.2-003"
/standard name="OTTHUMP00000002171"
/notes="match: proteins: Sw:Q13148 Sw:Q921F2 Tr:BAC27753"
/codon start=1
/product="TAR DNA binding protein"
/protein id="CA122098.1"
/db xref="GI:56204103"
/db xref="InterPro:IPR000504"
/translation="MSEYIRVTEDENDPEIPSEDDGTLLSTVTVAQPPGACGLRYR
NPSQMRGVLRGILHAPDAGMGLVYVNTPKDKRMDETDASSAVKRAVQK
TSLDVLGLPKFTTEQDLKFEYFTGEVLAVQVKDLKTHSGKGFVFTYEYQVK
VMSQRHMDICDCLPNKSKQSDPLRSKVFVGRCTEDMTEDLREFFFSQYGDVM
DVFIPKPRAFATVTDADQIAQSLGCEDLIKGISVHISNAEPKHSNRQLRSGRF
GNPFGNGGNGSGGAGLGNQGSNMGGMNGAFSINPAMAAQALQSSW
GMGMILASQNGSGGNNQNGMQRNPQAFSGSNYSYSGNSGNAATGWGSASNAG
SGSGFNGGFGSSMDSKSSGGM"
/join(18472..18598,19541..19677,20650..20683,21458..21552)
/gene="TARDBP"
/locus tag="RP4-635E18.2-007"
/join(18472..18598,19541..19677,20650..20683,21458..21552)
/gene="TARDBP"
/product="TAR DNA binding protein"
/notes="match: ESTs: Em:H26390.1"
/join(19546..19677,20412..20683,21458..21556,21899..22087)
/gene="TARDBP"
/locus tag="RP4-635E18.2-008"
/join(19546..19677,20412..20683,21458..21556,21899..22087)
/gene="TARDBP"
/locus tag="RP4-635E18.2-008"
/product="TAR DNA binding protein"
```

```
Query Match 87.2%; Score 21.8; DB 8; Length 112769;
Best Local Similarity 92.0%; Pred. No.16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
Db 25660 AAAAAACAATCAGGCCAGGCATGG 25636
```

```
RESULT 48
HS390C10/ c
LOCUS
DEFINITION
Human DNA sequence from clone CTA-390C10 on chromosome
22q11.21-12.1, complete sequence.
ACCESSION
AL008721
VERSION
AL008721.1 GI:3171883
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 114231)
Lloyd,D.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jun 2, 1998 this sequence version replaced gi:2598455.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep CTA-390C10 is
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBAC108L
```

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22

Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

```
FEATURES
            Location/Qualifiers
     1..114231
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="22"
        /map="g11.21-12.1"
        /clone="CTA-390C10"
        /clone_lib="CIT978SK-A1"
        join(complement(101..709),
        complement(AL022324.1:139775..140679))
        /locus tag="CTA-390C10.4-003"
        join(complement(101..709),
        complement(AL022324.1:139775..140679))
        /locus tag="CTA-390C10.4-003"
        complement(join(2233..2528,2650..2708))
        /locus tag="CTA-390C10.1-001"
        complement(join(2233..2528,2650..2708))
        /locus tag="CTA-390C10.1-001"
        join(12968..13156,20584..20712,22131..22273,24288..24553)
        /locus tag="CTA-390C10.2-001"
        /pseudo
        /locus tag="CTA-390C10.2-001"
        join(12968..13156,20584..20712,22131..22273,24288..24553)
        /pseudo
        /locus tag="CTA-390C10.6-006"
        42283..44821
        /locus tag="CTA-390C10.6-006"
        42283..44821
        /locus tag="CTA-390C10.6-006"
        join(74809..74942,75501..75780,76910..77156,85554..85605,
        86110..86290)
        /locus tag="CTA-390C10.3-003"
        join(74809..74942,75501..75780,76910..77156,85554..85605,
        86110..86290)
        /locus tag="CTA-390C10.3-003"
```

ORIGIN

```
Query Match 87.2%; Score 21.8; DB 8; Length 114231;
Best Local Similarity 92.0%; Pred. No.16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
Db 70553 AAAAAAAAATCTTGGCCAGGCATGG 70529
```

RESULT 49

```
AC078901/c
LOCUS
DEFINITION
Homo sapiens chromosome 13 clone RP11-474E23 map 13, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
ACCESSION
AC078901.4 GI:10280756
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 115909)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 13, clone RP11-474E23  
Unpublished  
2 (bases 1 to 115909)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,  
Boguslavsky,L., Boukagalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye-S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (10-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 115909)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bida,F., Boguslavsky,L.,  
Boukagalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 23, 2000 this sequence version replaced gi:9930750.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10224  
Center Clone name: 474\_E\_23  
----- Summary Statistics  
Sequencing vector: M13; M77815; 97% of reads  
Sequencing vector: Plasmid; n/a; 3% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 101354 bases at least Q40

Consensus quality: 109133 bases at least Q30  
Consensus quality: 111925 bases at least Q20  
Insert size: 121000; agarose-fp  
Insert size: 113209; sum-of-contigs  
Quality coverage: 5.0 in Q20 bases; agarose-fp  
Quality coverage: 5.3 in Q20 b.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 632: contig of 632 bp in length  
\* 633 732: gap of 100 bp  
\* 733 1761: contig of 1029 bp in length  
\* 1762 1861: gap of 100 bp  
\* 1866 2965: contig of 1104 bp in length  
\* 2966 3065: gap of 100 bp  
\* 3066 4724: contig of 1659 bp in length  
\* 4725 4824: gap of 100 bp  
\* 4825 6125: contig of 1301 bp in length  
\* 6126 6225: gap of 100 bp  
\* 6226 7471: contig of 1246 bp in length  
\* 7472 7571: gap of 100 bp  
\* 7572 8647: contig of 1076 bp in length  
\* 8648 8747: gap of 100 bp  
\* 8748 9829: contig of 1082 bp in length  
\* 9830 10967: contig of 1038 bp in length  
\* 10968 11067: gap of 100 bp  
\* 11068 12954: contig of 1887 bp in length  
\* 12955 13054: gap of 100 bp  
\* 13055 14354: contig of 1300 bp in length  
\* 14355 14454: gap of 100 bp  
\* 14455 15459: contig of 1005 bp in length  
\* 15460 15559: gap of 100 bp  
\* 15560 16757: contig of 1198 bp in length  
\* 16758 16857: gap of 100 bp  
\* 16858 19014: contig of 2157 bp in length  
\* 19015 19114: gap of 100 bp  
\* 19115 21051: contig of 1937 bp in length  
\* 21052 21151: gap of 100 bp  
\* 21152 23191: contig of 2040 bp in length  
\* 23192 23291: gap of 100 bp  
\* 23292 24664: contig of 1373 bp in length  
\* 24665 24764: gap of 100 bp  
\* 24765 27243: contig of 2479 bp in length  
\* 27244 27343: gap of 100 bp  
\* 27344 30774: contig of 3431 bp in length  
\* 30775 30874: gap of 100 bp  
\* 30875 32413: contig of 1539 bp in length  
\* 32414 32513: gap of 100 bp  
\* 32514 38423: contig of 5910 bp in length  
\* 38424 38523: gap of 100 bp  
\* 38524 44518: contig of 5955 bp in length  
\* 44519 44618: gap of 100 bp  
\* 44619 49673: contig of 5055 bp in length  
\* 49674 49773: gap of 100 bp  
\* 49774 57087: contig of 7314 bp in length  
\* 57088 57187: gap of 100 bp  
\* 57188 64533: contig of 7246 bp in length  
\* 64534 74157: contig of 9624 bp in length  
\* 74158 74257: gap of 100 bp  
\* 74258 89595: contig of 15338 bp in length  
\* 89596 89696: gap of 100 bp  
\* 89696 115909: contig of 26214 bp in length.  
Location/Qualifiers  
1. .115909  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

FEATURES  
source

```

/db xref="taxon:9606"
/clone="RP11-474E23"
/clone_lib="RPC1-11 Human Male BAC"
1. .632
/clone="assembly_fragment"
/clone_end:SP6
vector_side:left"
633. .732
/estimated_length=100
733. .1761
/notes="assembly_fragment"
1762. .1861
/estimated_length=100
1862. .2965
/notes="assembly_fragment"
2966. .3065
/estimated_length=100
3066. .4724
/notes="assembly_fragment"
4725. .4824
/estimated_length=100
4825. .6125
/notes="assembly_fragment"
6126. .6225
/estimated_length=100
6226. .7471
/notes="assembly_fragment"
7472. .7571
/estimated_length=100
7572. .8647
/notes="assembly_fragment"
8648. .8747
/estimated_length=100
8748. .9829
/notes="assembly_fragment"
9830. .9929
/estimated_length=100
9930. .10967
/notes="assembly_fragment"
10968. .11067
/estimated_length=100
11068. .12954
/notes="assembly_fragment"

Query Match      87.2%; Score 21.8; DB 14; Length 115909;
Best Local Similarity 92.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
Db 52366 AAAAAAAAAATCAGGCCAGGCATGG 52342

RESULT 50
CNS01DSM      117449 bp DNA linear PRI 26-APR-2001
LOCUS      Human chromosome 14 DNA sequence BAC C-2325P2 of library CalTech-D
DEFINITION      from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION      AL121820
VERSION      AL121820.4 GI:13872724
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Hominidae; Homo.
REFERENCE      1 (bases 1 to 117449)
AUTHORS      Heilig, R., Petit, J.-L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.

Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 117449)
Genoscope.
Direct Submission
Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr]
- Web : www.genoscope.cns.fr)
On Apr 27, 2001 this sequence version replaced gi:8176917.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-1082A1 (AC-AL13313)
Downstream BAC (overlapping the SP6 end) : R-108B17 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.50x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 : 1
30 - 39 : 24
40 - 49 : 1973
50 - 59 : 5247
60 - 69 : 6446
70 - 79 : 14771
80 - 89 : 39167
90 - 99 : 49820
-----
Percentage of bases with a quality value >= 40 : 99 %

FEATURES
source
Location/Qualifiers
1. .117449
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="C-2325P2"
/clone_lib="CalTech-D"
81976. .82149
/notes="matching EMBL:W61011
RHdb:RH103969
dbSTS:STS71498
Identified using the e-PCR software (G. Schuler)"
82052. .82258
/notes="matching EMBL:G37269
RHdb:RH100968
RHdb:RH81746
dbSTS:STS58098
Identified using the e-PCR software (G. Schuler)"

ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 117449;
Best Local Similarity 92.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
Db 52050 AAAAAAAAAATCAGGCCAGGCATGG 52074

Search completed: February 3, 2006, 21:25:46
Job time : 877 secs

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-1  
Perfect score: 25  
Sequence: 1 aaaaaaaatcacggcagcgatgg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Maximum first 100%  
Listing first 150 summaries

Database : N.Geneseq\_21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.8	87.2	341	AAK74376	Aak74376 Human imm
2	21.8	87.2	464	ADDD35001	Add35001 Mouse mit
C 3	21.8	87.2	2825	ADL06341	Adl06341 cDNA enco
C 4	21.8	87.2	3067	AAH19210	Aah19210 Human sec
C 5	21.8	87.2	3067	ADD67600	Add67600 Human Ly1
C 6	21.8	87.2	3105	AAH19178	Aah19178 Human sec
7	21.8	87.2	3288	ACC72030	Acc72030 BC00571 g
8	21.8	87.2	5017	AAK69927	Aak69927 Human imm
9	21.8	87.2	5017	AAK69928	Aak69928 Human imm
C 10	21.8	87.2	98642	ACN44584	Acn44584 Mouse gen
C 11	21.8	87.2	110000	ACN45090_0	Acn45090 Human gen
12	20.8	83.2	25574	AAK79671	Aak79671 Human imm
13	20.8	83.2	25574	AAK83760	Aak83760 Human imm
14	20.8	83.2	25574	AAK05619	Aal05619 Human rep
15	20.8	83.2	25576	AAK85305	Aak85305 Human imm
16	20.8	83.2	25576	AAK79669	Aak79669 Human imm
17	20.8	83.2	25576	AAK83758	Aak83758 Human imm
18	20.8	83.2	25576	AAK05618	Aal05618 Human rep
C 19	20.8	83.2	110000	11 ACN44934_1	Continuation (2 of

Continuation (4 of	ADZ46976_3	14	83.2	110000	20.8
Adl18927 Human sof	ADQ18927	12	81.6	210204	20.4
Aac17433 Human sec	AAC17433	3	80.8	138	20.2
Aac14155 Human sec	AAC14155	3	80.8	206	20.2
Ac156670 Human col	ACL56670	302	80.8	302	20.2
Aak74375 Human imm	AAK74375	339	80.8	339	20.2
Aas37660 Novel hum	AAS37660	344	80.8	344	20.2
Abk95389 Human ret	ABK95389	384	80.8	384	20.2
Aak83882 Human imm	AAK83882	392	80.8	392	20.2
Aai187906 Human pol	AAI187906	403	80.8	403	20.2
Ab182000 Human ova	ABL82000	465	80.8	465	20.2
Acf85697 Human sir	ACF85697	13	80.8	465	20.2
Abn61750 Human can	ABN61750	480	80.8	480	20.2
Ac160088 Human col	ACL60088	14	80.8	480	20.2
Abas59354 Human foe	ABA59354	484	80.8	484	20.2
Aai39155 Probe #78	AAI39155	484	80.8	484	20.2
Aak33368 Human bon	AAK33368	484	80.8	484	20.2
Aak07570 Human bra	AAK07570	484	80.8	484	20.2
Abas33124 Human liv	ABS33124	484	80.8	484	20.2
Abas08206 Human gen	ABS08206	484	80.8	484	20.2
Abn64674 Human can	ABN64674	526	80.8	526	20.2
Aaa35120 Human ade	AAA35120	666	80.8	666	20.2
Aaf21242 Human low	AAF21242	3	80.8	666	20.2
Abz66936 Human nuc	ABZ66936	10	80.8	666	20.2
Abd20785 Human pul	ABD20785	11	80.8	666	20.2
Aas28505 Genomic s	AAS28505	1131	80.8	1131	20.2
Aas29996 Human lun	AAS29996	5	80.8	1131	20.2
Adb33333 Human nov	ADB33333	10	80.8	1131	20.2
Adg41701 Human res	ADG41701	11	80.8	1131	20.2
Adi97475 Human res	ADI97475	11	80.8	1131	20.2
Abq75320 Human lun	ABQ75320	6	80.8	1264	20.2
Abz58228 Human G-p	ABZ58228	8	80.8	1271	20.2
Acn37887 Tumour-as	ACN37887	13	80.8	1935	20.2
Aak85252 Human imm	AAK85252	4	80.8	2713	20.2
Abz74454 Secreted	ABZ74454	8	80.8	3310	20.2
Ada98878 Human sec	ADA98878	8	80.8	3310	20.2
Adc20907 Human sec	ADC20907	10	80.8	3310	20.2
Abt16993 Human sec	ABT16993	10	80.8	3310	20.2
Abz68001 Human sec	ABZ68001	10	80.8	3310	20.2
Adg64697 Novel hum	ADG64697	12	80.8	3706	20.2
Abz81729 Benign pr	ABZ81729	10	80.8	3943	20.2
Abz68715 Human imm	ABZ68715	4	80.8	4013	20.2
Aav99094 DNA methy	AAV99094	2	80.8	4459	20.2
Adg63025 Novel hum	ADG63025	12	80.8	5003	20.2
Aas28641 Genomic s	AAS28641	6	80.8	6186	20.2
Adg41837 Human res	ADG41837	10	80.8	6186	20.2
Adi97611 Human res	ADI97611	11	80.8	6186	20.2
Aas28642 Genomic s	AAS28642	4	80.8	6191	20.2
Aas28643 Genomic s	AAS28643	4	80.8	6191	20.2
Adg41838 Human res	ADG41838	10	80.8	6191	20.2
Adg41839 Human res	ADG41839	10	80.8	6191	20.2
Adi97613 Human res	ADI97613	11	80.8	6191	20.2
Adi97612 Human res	ADI97612	11	80.8	6191	20.2
Aal44086 Human tra	AAL44086	6	80.8	6339	20.2
Aak70928 Human imm	AAK70928	4	80.8	6743	20.2
Aal05842 Human rep	AAL05842	4	80.8	6743	20.2
Ab198406 Human tes	ABL98406	4	80.8	6743	20.2
Abk46568 DNA enco	ABK46568	6	80.8	7426	20.2
Aak83896 Human imm	AAK83896	4	80.8	7794	20.2
Aas28608 Genomic s	AAS28608	4	80.8	7794	20.2
Adg41804 Human res	ADG41804	10	80.8	7794	20.2
Adi97578 Human res	ADI97578	11	80.8	7794	20.2
Aal51033 Human Ras	AAL51033	8	80.8	9454	20.2
Aas28920 Human imm	AAS28920	4	80.8	10236	20.2
Aas28880 Human imm	AAS28880	4	80.8	10236	20.2
Aak88907 Human imm	AAK88907	4	80.8	10236	20.2
Aak80678 Human imm	AAK80678	4	80.8	10236	20.2
Aak68717 Human imm	AAK68717	4	80.8	10236	20.2
Abas06792 Human gen	ABA06792	6	80.8	10236	20.2
Abv84129 Human pol	ABV84129	6	80.8	10236	20.2
Adb31721 Human nov	ADB31721	10	80.8	10236	20.2
Adb31761 Human neu	ADB31761	10	80.8	10236	20.2
Aaf97872 Human neu	AAF97872	5	80.8	12312	20.2

```
C 93 20.2 80.8 13224 4 AAS41751
C 94 20.2 80.8 13224 4 ABA06811
C 95 20.2 80.8 13224 6 ABV84148
C 96 20.2 80.8 15000 6 ABL62905
C 97 20.2 80.8 15866 4 AAK90892
C 98 20.2 80.8 16500 12 ADP68405
C 99 20.2 80.8 17425 2 AAX28366
C 100 20.2 80.8 19301 3 AAX35123
C 101 20.2 80.8 19301 3 AAF21245
C 102 20.2 80.8 19301 10 ABZ96939
C 103 20.2 80.8 19301 11 ABD20788
C 104 20.2 80.8 19440 2 AAV99129
C 105 20.2 80.8 19802 4 AAK90894
C 106 20.2 80.8 19802 4 AAL04554
C 107 20.2 80.8 19802 4 ABL97477
C 108 20.2 80.8 21403 4 AAK90893
C 109 20.2 80.8 21403 4 AAL04555
C 110 20.2 80.8 21403 4 ABL97478
C 111 20.2 80.8 23763 4 AAK90895
C 112 20.2 80.8 23763 4 AAL04558
C 113 20.2 80.8 23763 4 ABL97481
C 114 20.2 80.8 24977 4 AAL03263
C 115 20.2 80.8 24977 4 AAL05788
C 116 20.2 80.8 24977 4 ABL98352
C 117 20.2 80.8 24983 4 AAL05787
C 118 20.2 80.8 24983 4 AAL03282
C 119 20.2 80.8 24983 4 ABL98351
C 120 20.2 80.8 26354 12 ADQ97798
C 121 20.2 80.8 27062 4 AAK71622
C 122 20.2 80.8 27062 4 AAL05346
C 123 20.2 80.8 27062 4 ABL98215
C 124 20.2 80.8 27990 11 AGL44366
C 125 20.2 80.8 28449 12 ADQ97240
C 126 20.2 80.8 29329 4 AAK78512
C 127 20.2 80.8 29329 4 AAK70791
C 128 20.2 80.8 29329 5 ABA18026
C 129 20.2 80.8 29329 5 ABA20511
C 130 20.2 80.8 30573 4 AAL04181
C 131 20.2 80.8 32167 4 AAL05509
C 132 20.2 80.8 32191 4 AAL199472
C 133 20.2 80.8 32191 4 AAI64117
C 134 20.2 80.8 32191 12 ADF71715
C 135 20.2 80.8 32195 4 AAL04180
C 136 20.2 80.8 32622 11 AGL43890
C 137 20.2 80.8 43256 12 ADQ97736
C 138 20.2 80.8 44840 4 AAK71803
C 139 20.2 80.8 44978 13 ADS36458
C 140 20.2 80.8 45546 2 AAX23520
C 141 20.2 80.8 47243 11 AGL44690
C 142 20.2 80.8 52616 4 AAK70459
C 143 20.2 80.8 52616 4 AAK78930
C 144 20.2 80.8 52710 13 ABD33436
C 145 20.2 80.8 57038 11 ADP65343
C 146 20.2 80.8 57130 8 ABO77243
C 147 20.2 80.8 65559 11 AGL44408
C 148 20.2 80.8 65608 6 ABL62910
C 149 20.2 80.8 65608 6 ABL64414
C 150 20.2 80.8 65608 6 ABL67668
```

## ALIGNMENTS

```
RESULT 1
ID AAK74376
XX AAK74376 standard; DNA; 341 BP.
AC AAK74376;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:29188.
XX
```

```
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
PN WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2002; 2000US-0179065P.
XX 04-FEB-2002; 2000US-0180628P.
XX 24-FEB-2002; 2000US-0184664P.
XX 02-MAR-2002; 2000US-0186350P.
XX 16-MAR-2002; 2000US-0189874P.
XX 17-MAR-2002; 2000US-0190076P.
XX 18-APR-2002; 2000US-0198123P.
XX 19-MAY-2002; 2000US-0205515P.
XX 07-JUN-2002; 2000US-0209467P.
XX 28-JUN-2002; 2000US-0214886P.
XX 30-JUN-2002; 2000US-0215135P.
XX 07-JUL-2002; 2000US-0216647P.
XX 07-JUL-2002; 2000US-0216880P.
XX 11-JUL-2002; 2000US-0217487P.
XX 11-JUL-2002; 2000US-0217496P.
XX 14-JUL-2002; 2000US-0218290P.
XX 26-JUL-2002; 2000US-0220963P.
XX 26-JUL-2002; 2000US-0220964P.
XX 14-AUG-2002; 2000US-0224518P.
XX 14-AUG-2002; 2000US-0224519P.
XX 14-AUG-2002; 2000US-0225213P.
XX 14-AUG-2002; 2000US-0225214P.
XX 14-AUG-2002; 2000US-0225266P.
XX 14-AUG-2002; 2000US-0225267P.
XX 14-AUG-2002; 2000US-0225268P.
XX 14-AUG-2002; 2000US-0225270P.
XX 14-AUG-2002; 2000US-0225447P.
XX 14-AUG-2002; 2000US-0225757P.
XX 14-AUG-2002; 2000US-0225758P.
XX 14-AUG-2002; 2000US-0225759P.
XX 18-AUG-2002; 2000US-0226279P.
XX 22-AUG-2002; 2000US-0226681P.
XX 22-AUG-2002; 2000US-0226686P.
XX 22-AUG-2002; 2000US-0227182P.
XX 23-AUG-2002; 2000US-0227009P.
XX 30-AUG-2002; 2000US-0228924P.
XX 01-SEP-2002; 2000US-0229287P.
XX 01-SEP-2002; 2000US-0229343P.
XX 01-SEP-2002; 2000US-0229344P.
XX 01-SEP-2002; 2000US-0229345P.
XX 05-SEP-2002; 2000US-0229509P.
XX 05-SEP-2002; 2000US-0229513P.
XX 06-SEP-2002; 2000US-0230437P.
XX 06-SEP-2002; 2000US-0230438P.
XX 08-SEP-2002; 2000US-0231242P.
XX 08-SEP-2002; 2000US-0231243P.
XX 08-SEP-2002; 2000US-0231244P.
XX 08-SEP-2002; 2000US-0231413P.
XX 08-SEP-2002; 2000US-0231414P.
XX 08-SEP-2002; 2000US-0232080P.
XX 08-SEP-2002; 2000US-0232081P.
XX 12-SEP-2002; 2000US-0231968P.
XX 14-SEP-2002; 2000US-0232397P.
XX 14-SEP-2002; 2000US-0232398P.
XX 14-SEP-2002; 2000US-0232399P.
XX 14-SEP-2002; 2000US-0232400P.
XX 14-SEP-2002; 2000US-0232401P.
XX 14-SEP-2002; 2000US-0233063P.
XX 14-SEP-2002; 2000US-0233064P.
XX 14-SEP-2002; 2000US-0233065P.
XX 21-SEP-2002; 2000US-0234223P.
XX 21-SEP-2002; 2000US-0234274P.
```

```
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234984P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 29188; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX SQ Sequence 341 BP; 133 A; 49 C; 77 G; 82 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 4; Length 341;
Best Local Similarity 92.0%; Pred. No. 7.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 236 AAAAAAAAAAACAGCGCCAGGCATGG 260

RESULT 2
ADD35001
ID ADD35001 standard; DNA; 464 BP.
XX AC ADD35001;
XX 15-JAN-2004 (first entry)
XX Mouse mitochondrial DNA sequence SEQ ID NO:2781.
XX db; mouse; array; mitochondrial; hybridisation; energy-metabolism;
XX mitochondrial disease; oxidative phosphorylation dysfunction;
XX oxidative stress; apoptosis; aging.
XX Mus musculus.
XX WO2003020220-A2.
XX 13-MAR-2003.
XX 30-AUG-2002; 2002WO-US027886.
XX 30-AUG-2001; 2001US-0316323P.
XX 31-AUG-2001; 2001CA-02356540.
XX (UYEM-) UNIV EMORY.
XX Wallace DC, Levy S, Kerstann K, Procaccio V;
XX WPI; 2003-300821/29.
```

PT Array containing probes for genes involved in mitochondrial biology,  
 PT useful for determining mitochondrial biology gene expression profiles for  
 PT use in diagnosing pathologies and identifying biochemical pathways.

XX Claim 2; SEQ ID NO 2781; 201pp; English.

XX The invention relates to a novel array comprising at least two isolated  
 CC nucleotide molecules, each molecule having a sequence capable of uniquely  
 CC hybridising to a nucleic acid molecule which is an expression product of  
 CC a gene involved in mitochondrial biology. The array comprises two or more  
 CC isolated nucleic acid molecules or spots, each molecule having a sequence  
 CC chosen from sequence of 994 human probes and 2046 mouse probes. An array  
 CC of the invention is useful for determining an expression profile of a  
 CC mouse or human sample containing nucleic acid, by contacting the array  
 CC with the sample under conditions allowing selective hybridisation, and  
 CC measuring hybridisation of nucleic acid in the sample to the array to  
 CC produce an expression profile. The array is also useful for determining  
 CC an expression profile of a first labelled sample containing nucleic acid  
 CC relative to a second, differently labelled sample containing nucleic  
 CC acid. The second sample is a reference or a standard. An array is useful  
 CC for determining an expression profile diagnostic of an energy-metabolism-  
 CC related physiological condition. An array of the invention is useful for  
 CC determining mitochondrial biology gene expression profiles of organisms,  
 CC such as human, mice and closely related species, tissue and organs of  
 CC such organisms, which are useful for determining expression profiles  
 CC diagnostic of energy metabolism-related physiological conditions.  
 CC diagnosing such physiological conditions, identifying biochemical  
 CC pathways, genes, and mutations involved in such physiological conditions,  
 CC identifying therapeutic agents useful for preventing and/or treating such  
 CC physiological conditions, evaluating and/or monitoring the efficacy of  
 CC such therapies, and creating and identifying animal models of human  
 CC energy metabolism-related physiological conditions. An array is also  
 CC useful for defining expression signatures or profiles for mitochondrial  
 CC diseases, as well as distinguishing clinical disorders that result from  
 CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,  
 CC apoptosis and aging. An array of the invention contains probes of genes  
 CC not previously recognised to participate in mitochondrial biology. The  
 CC sequences shown in ADP33224-ADP35260 represent murine mitochondrial DNA  
 CC clones used to make the probes of the invention. Some sequences are not  
 CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,  
 CC 1906, 2408 and 2643.

XX Sequence 464 BP; 180 A; 91 C; 80 G; 111 T; 0 U; 2 Other;

Query Match 87.2%; Score 21.8; DB 10; Length 464;  
 Best Local Similarity 92.0%; Pred. No. 8;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 255 AAAAAAAAAATCTTGGCCAGGCATGG 279

RESULT 3

ADL06341/c

ID ADL06341 standard; cDNA; 2825 BP.

AC ADL06341;

DT 06-MAY-2004 (first entry)

DE cDNA encoding human protein-13.2, a site-specific recombinase.

XX Human; protein-13.2; site-specific recombinase;  
 KW growth development disorder; tumour; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 1016..1378

FT CDS /\*tag= a

FT /\*product= "Protein-13.2"

XX

PN CN1393548-A.

XX 29-JAN-2003.

PD 29-JAN-2003.

PF 29-JUN-2001; 2001CN-001131178.

XX 29-JUN-2001; 2001CN-001131178.

PR (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.

PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.

XX Mao Y, Xie Y;

PI WPI; 2003-422181/40.

XX P-PSDB; ADL06342.

XX Polypeptide-human protein-13.2 containing site specific recombinase

PT characteristic sequence fragment and polynucleotide for coding it.

PT Claim 6; SEQ ID NO 1; 32pp; Chinese.

XX The present invention relates to the isolation of human protein-13.2

CC containing a site-specific recombinase characteristic sequence fragment,

CC and the polynucleotide sequence encoding it. Also disclosed is a process

CC for preparing the polypeptide by a DNA recombination technique and

CC application of the polypeptide and polynucleotide in treating diseases

CC such as growth development disorders and tumours. The present sequence

CC encodes human protein-13.2.

XX Sequence 2825 BP; 799 A; 556 C; 707 G; 763 T; 0 U; 0 Other;

SQ Query Match 87.2%; Score 21.8; DB 10; Length 2825;

Best Local Similarity 92.0%; Pred. No. 11;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 291 AAAAAAAAAATTAAGCCAGGCATGG 267

RESULT 4

AAH19210/c

ID AAH19210 standard; cDNA; 3067 BP.

AC AAH19210;

XX 17-JUL-2001 (first entry)

DT Human secreted protein-encoding gene 9 cDNA clone HTOHB55, SEQ ID NO:51.

DE Human; secreted protein; proliferative disorder; cancer;

XX foetal abnormality; developmental abnormality; haematopoietic disorder;

XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

XX inflammation; allergy; neurological disorder; Alzheimer's disease;

XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;

XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

XX cardiovascular disorder; angioinetic disorder; kidney disorder;

XX gastrointestinal disorder; pregnancy-related disorder; tumour;

XX endocrine disorder; infection; wound healing; vulnerability; cell culture;

XX chemotaxis; food additive; binding partner identification; ss.

XX Homo sapiens.

OS WO200132910-A2.

XX 10-MAY-2001.

PD 25-OCT-2000; 2000WO-US029362.

XX 29-OCT-1999; 99US-0162240P.

PR 30-JUN-2000; 2000US-0215131P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA

PI Ruben SM, Komatsoulis GA, Birse CE, Ni J, Soppet DR;  
XX WPI; 2001-335835/35.  
DR N-PSDB; AAG62196.  
XX  
XX Novel 27 isolated human secreted proteins and polynucleotides encoding  
PT them useful for treating, diagnosing, preventing Alzheimer's disease,  
PT Parkinson's disease, AIDS, rheumatoid arthritis, asthma.  
XX  
XX Claim 4; Page 469-470; 594pp; English.  
XX  
XX AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted  
CC protein genes, and AAG62156-AAG62235 represent the proteins they encode.  
CC AAG62236-AAG62293 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin  
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiotensin-related disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention  
XX  
SQ Sequence 3067 BP; 672 A; 826 C; 783 G; 786 T; 0 U; 0 Other;  
  
Query Match 87.2%; Score 21.8; DB 4; Length 3067;  
Best Local Similarity 92.0%; Pred. No. 11;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 1184 AAAAAATAATCAGCGCCAGGCATGG 1160  
  
RESULT 5  
ADD67600/c  
ID ADD67600 standard; cDNA; 3067 BP.  
XX  
AC ADD67600;  
XX  
XX 15-JAN-2004 (first entry)  
DT  
XX  
DE Human Ly1485P partial cDNA SEQ ID NO:77.  
XX  
XX haematological malignancy; immunoconjugate; cytostatic; immunostimulant;  
KW vaccine; immunotherapy; cancer; multiple myeloma cell;  
KW chronic lymphocytic leukaemia; B cell leukaemia; lymphoma; anti-cancer;  
KW human; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO2003062401-A2.  
PN  
XX 31-JUL-2003.  
PD  
XX

PF 22-JAN-2003; 2003WO-US002353.  
XX  
PR 22-JAN-2002; 2002US-00057475.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Gaiger A, Algate PA, Mannion J, Clapper JD, Wang A, Ordenez N;  
PI Carter L, McNeill PD;  
XX  
XX WPI; 2003-598749/56.  
DR  
XX  
XX New hematological malignancy-related genes and polypeptides, useful for  
PT screening anti-cancer agents, and generating antibodies or  
PT immunoconjugates for treating e.g. multiple myeloma cell or chronic  
PT lymphocytic leukemia.  
XX  
XX Claim 1; SEQ ID NO 77; 307pp; English.  
XX  
XX The present invention describes an isolated polynucleotide (I), which is  
CC overexpressed in haematological malignancies, and which encodes a  
CC polypeptide or an immunogenic fragment of the polypeptide. Also  
CC described: (1) an isolated polypeptide; (2) an expression vector  
CC comprising (1) operably linked to an expression control sequence; (3) a  
CC host cell comprising an expression vector; (4) an isolated antibody that  
CC specifically binds to the polypeptide or its immunogenic fragment; and  
CC (5) immunoconjugates comprising the antibody above, or an antibody that  
CC specifically binds to a polypeptide, or its immunogenic fragment, encoded  
CC by (1). (I) has cytostatic and immunostimulant activities, and can be  
CC used in vaccines and immunotherapy. The immunoconjugates are useful in  
CC the manufacture of a medicament, particularly as active ingredients in a  
CC composition for treating cancer, e.g. multiple myeloma cell, chronic  
CC lymphocytic leukaemia, B cell leukaemias, or lymphomas in humans, sheep,  
CC primates, goats, bovines, equines, porcines, lupines, canines or felines.  
CC The polynucleotide (I) or polypeptide can be used for screening anti-  
CC cancer agents, and generating antibodies or immunoconjugates for treating  
CC or preventing the above-mentioned diseases. The polynucleotide,  
CC polypeptide or antibody can be used for detecting, diagnosing or  
CC prognosticating the haematological malignancies described above. The  
CC present sequence is used in the exemplification of the present invention.  
XX  
SQ Sequence 3067 BP; 672 A; 826 C; 783 G; 786 T; 0 U; 0 Other;  
  
Query Match 87.2%; Score 21.8; DB 10; Length 3067;  
Best Local Similarity 92.0%; Pred. No. 11;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 1184 AAAAAATAATCAGCGCCAGGCATGG 1160  
  
RESULT 6  
AAH19178/c  
ID AAH19178 standard; cDNA; 3105 BP.  
XX  
AC AAH19178;  
XX  
XX 17-JUL-2001 (first entry)  
DT  
XX  
DE Human secreted protein-encoding gene 9 cDNA clone HTOHB55, SEQ ID NO:19.  
XX  
XX Human secreted protein; proliferative disorder; cancer;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiotensin-related disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; tumour;  
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
KW chemotaxis; food additive; binding partner identification; ss.  
XX  
XX Homo sapiens.  
OS

```
XX WO200132910-A2.
PN 10-MAY-2001.
XX
XX 25-OCT-2000; 2000WO-US029362.
XX
XX 29-OCT-1999; 99US-0162240P.
PR 30-JUN-2000; 2000US-0215131P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Birse CE, Ni J, Soppet DR;
PI WPI; 2001-335835/35.
XX N-PSDB; AAG62164.
XX
XX Novel 27 isolated human secreted proteins and polynucleotides encoding
PT them useful for treating, diagnosing, preventing Alzheimer's disease,
PT Parkinson's disease, AIDS, rheumatoid arthritis, asthma.
XX
XX Claim 4; Page 449-450; 594pp; English.
XX
XX AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted
CC protein genes, and AAG62156-AAG62235 represent the proteins they encode.
CC AAG62236-AAG62293 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention
XX
XX SQ Sequence 3105 BP; 678 A; 833 C; 795 G; 791 T; 0 U; 8 Other;
Query Match 87.2%; Score 21.8; DB 4; Length 3105;
Best Local Similarity 92.0%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 1177 AAAAAATATCAGCGCCAGGCACGG 1153
XX
XX RESULT 7
ACCT72030
ID ACC72030 standard; DNA; 3288 BP.
XX
XX ACC72030;
XX
XX 08-JUL-2003 (first entry)
XX
XX BCU0571 gene #SEQ ID 37.
XX
```

```
XX Breast cancer; cytostatic; gene therapy; antisense therapy; regulated;
KW drug discovery; clinical medicine; forensic medicine; gene;
KW chromosome 8q24.12; ds.
XX
XX Homo sapiens.
XX
XX WO2003029421-A2.
XX
XX 10-APR-2003.
XX
XX 02-OCT-2002; 2002WO-US031287.
XX
XX 03-OCT-2001; 2001US-0326526P.
PR 14-MAY-2002; 2002US-00144194.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Li X, Fan W, Kovacs KF, Jay G;
PI WPI; 2003-381623/36.
XX P-PSDB; ABR58296.
XX
XX New isolated human differentially-regulated breast cancer polynucleotide
PT and polypeptide, useful for diagnosing, staging, prognosticating,
PT preventing and/or treating diseases and conditions relating to breast
PT cancer.
XX
XX Claim 2; SEQ ID NO 37; 127pp + Sequence Listing; English.
XX
XX The invention relates to isolated polynucleotides which are
CC differentially-regulated in breast cancer. The methods and compositions
CC of the present invention are useful for detecting, diagnosing, staging,
CC monitoring, prognosticating, preventing and/or treating diseases and
CC conditions relating to breast cancer, and may be used in gene therapy or
CC antisense therapy. They can also be used in research, drug discovery,
CC clinical medicine and forensic medicine. Sequences given in records
CC ACC72012-ACC72074 represent polynucleotides of the invention that are
CC differentially-regulated in breast cancer. NOTE: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 3288 BP; 967 A; 582 C; 753 G; 986 T; 0 U; 0 Other;
Query Match 87.2%; Score 21.8; DB 8; Length 3288;
Best Local Similarity 92.0%; Pred. No. 11;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 2979 AAAAAAATCTTGGCCAGGCATGG 3003
XX
XX RESULT 8
AAK69927
ID AAK69927 standard; DNA; 5017 BP.
XX
XX AAK69927;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24739.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
```



PF 17-JAN-2001; 2001WO-US001354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180828P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 13-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 23-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-02559678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides.

PT useful for preventing, diagnosing and/or treating cancers and metastasis.

XX Disclosure, SEQ ID NO 24739; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention

XX SQ Sequence 5017 BP; 1338 A; 963 C; 1013 G; 1703 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 4; Length 5017;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

||||||| |||||||

Db 3079 AAAAAAAAAAACGCCAGGCATGG 3103

RESULT 9

AAK69928

ID AAK69928 standard; DNA; 5017 BP.

AC AAK69928;

XX 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24740.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229144P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.



WP Sequence split into 4 fragments LOCUS ACN45090 Accession ACN45090  
WP Fragment Name Begin End  
WP ACN45090\_0 1 110000  
WP ACN45090\_1 100001 210000  
WP ACN45090\_2 200001 310000  
WP ACN45090\_3 300001 350764  
ID ACN45090 standard; DNA; 350764 BP.  
XX  
AC ACN45090;

XX  
XX 18-NOV-2004 (first entry)  
XX Human genomic sequence hCG22125.  
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX Homo sapiens.

XX WO2003073826-A2.  
XX 12-SEP-2003.  
XX 28-FEB-2003; 2003WO-US006235.  
XX 01-MAR-2002; 2002US-00087192.  
XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;  
XX WPI; 2003-328604/31.  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX Claim 1; SEQ ID NO 1864; Opp; English.  
XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published

SQ Sequence 350764 BP; 99569 A; 63584 C; 68015 G; 119596 T; 0 U; 0 Other;  
Query Match 87.2%; Score 21.8; DB 11; Length 110000;  
Best Local Similarity 92.0%; Pred. No. 20;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
Db 23134 AAAAAAAAAAAGCCGCGCATGG 23110

RESULT 12  
AAK79671  
ID AAK79671 standard; DNA; 25574 BP.  
XX  
AC AAK79671;  
XX  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34483.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209457P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226727P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
XX 14-SEP-2000; 2000US-0232401P.  
XX 14-SEP-2000; 2000US-0233063P.  
XX 14-SEP-2000; 2000US-0233064P.  
XX 14-SEP-2000; 2000US-0233065P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.

```

PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251478P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 11-DEC-2000; 2000US-0254097P.

PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 34483; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
XX SQ Sequence 25574 BP; 7154 A; 6018 C; 6058 G; 6344 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 4; Length 25574;
Best Local Similarity 91.7%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATG 24
   |||||
Db 12227 AAAAAAAAAATCAGCGCCAGCGCATG 12250

RESULT 13
AAK83760
ID AAK83760 standard; DNA; 25574 BP.
XX AAK83760;
AC
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38572.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
```

PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0231968P.  
PR 12-SEP-2000; 2000US-0231969P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241828P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
PT Disclosure; SEQ ID NO 38572; 307lpp + Sequence Listing; English.  
PS AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 25574 BP; 7154 A; 6018 C; 6058 G; 6344 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 4; Length 25574;  
Best Local Similarity 91.7%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCCGAGCATG 24  
|||||||  
Db 12227 AAAAAAAAAATCAGCCGAGCATG 12250

## RESULT 14

AAL05619

ID AAL05619 standard; DNA; 25574 BP.

XX AC AAL05619;

XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen DNA SEQ ID NO: 8307.

XX KW Human; reproductive system related antigen; reproductive system disorder;  
XX KW cancer; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217496P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 14-AUG-2000; 2000US-0225759P.

XX PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234423P.  
PR 21-SEP-2000; 2000US-0234424P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 29-SEP-2000; 2000US-0236371P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241321P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.



PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259578P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
XX WPI; 2001-465570/50.  
DR  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX Disclosure; SEQ ID NO 8307; 1297bp + Sequence Listing; English.  
PS  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
XX Sequence 25574 BP; 7154 A; 6018 C; 6058 G; 6344 T; 0 U; 0 Other;  
SQ  
Query Match 83.2%; Score 20.8; DB 4; Length 25574;  
Best Local Similarity 91.7%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCAGCGATG 24  
|||||||  
Db 12227 AAAAAAAAAATCAAGCGCAGCGTG 12250  
RESULT 15  
AAK85305  
ID AAK85305 standard; DNA; 25576 BP.  
XX  
AC AAK85305;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40117.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
KW

XX  
OS Homo sapiens.  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.

PR	26-SEP-2000;	2000US-0235484P.	PA	(HUMA-) HUMAN GENOME SCI INC.
PR	27-SEP-2000;	2000US-0235834P.	XX	
PR	27-SEP-2000;	2000US-0235836P.	PI	Rosen CA, Barash SC, Ruben SM;
PR	29-SEP-2000;	2000US-0236327P.	XX	
PR	29-SEP-2000;	2000US-0236367P.	DR	WPI; 2001-483426/52.
PR	29-SEP-2000;	2000US-0236368P.	XX	
PR	29-SEP-2000;	2000US-0236369P.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PR	29-SEP-2000;	2000US-0236370P.	PT	useful for preventing, diagnosing and/or treating cancers and metastasis.
PR	02-OCT-2000;	2000US-0236802P.	XX	
PR	02-OCT-2000;	2000US-0237037P.	PS	Disclosure; SEQ ID NO 40117; 3071pp + Sequence Listing; English.
PR	02-OCT-2000;	2000US-0237038P.	XX	
PR	02-OCT-2000;	2000US-0237039P.	CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
PR	02-OCT-2000;	2000US-0237040P.	CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
PR	13-OCT-2000;	2000US-0239935P.	CC	activity, and can be used in gene therapy and vaccine production. (I)
PR	13-OCT-2000;	2000US-0239937P.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and
PR	20-OCT-2000;	2000US-0240960P.	CC	treatment of diseases associated with inappropriate (I) expression. For
PR	20-OCT-2000;	2000US-0241221P.	CC	example, they may be used to treat disorders associated with decreased
PR	20-OCT-2000;	2000US-0241785P.	CC	expression by rectifying mutations or deletions in a patient's genome
PR	20-OCT-2000;	2000US-0241786P.	CC	that affect the activity of (I) by expressing inactive proteins or to
PR	20-OCT-2000;	2000US-0241787P.	CC	supplement the patients own production of (I). Additionally, (I)
PR	20-OCT-2000;	2000US-0241808P.	CC	polynucleotides may be used to produce the secreted (I), by inserting the
PR	20-OCT-2000;	2000US-0241809P.	CC	nucleic acids into a host cell and culturing the cell to express the
PR	01-NOV-2000;	2000US-0246176P.	CC	protein. (I) proteins and polynucleotides may be used to prevent,
PR	01-NOV-2000;	2000US-0246177P.	CC	diagnose and treat immune/haematopoietic-related diseases, especially
PR	08-NOV-2000;	2000US-0246474P.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703
PR	08-NOV-2000;	2000US-0246475P.	CC	to AAK87694 represent human immune/haematopoietic antigen genomic
PR	08-NOV-2000;	2000US-0246476P.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169
PR	08-NOV-2000;	2000US-0246477P.	CC	represent sequences used in the exemplification of the present invention
XX	Sequence 25576 BP; 7153 A; 6020 C; 6060 G; 6343 T; 0 U; 0 Other;		XX	
Query Match 83.2%; Score 20.8; DB 4; Length 25576;				
Best Local Similarity 91.7%; Pred. No. 44;				
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1	AAAAAAAAATCAGCGCCAGCGATG 24		
Db	12225	AAAAAAAAATCAAGCGCCAGCGTG 12248		
RESULT 16				
AAK79669				
ID	AAK79669	standard; DNA; 25576 BP.		
AC	AAK79669;			
XX				
DT	07-NOV-2001	(first entry)		
XX				
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34481.			
XX				
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;			
KW	cytostatic; gene therapy; vaccine; immune/haematopoietic antigen; cancer;			
XX				
OS	Homo sapiens.			
XX				
PN	WO200157182-A2.			
XX				
PD	09-AUG-2001.			
XX				
PF	17-JAN-2001; 2001WO-US001354.			
XX				
PR	31-JAN-2000; 2000US-0179065P.			
PR	04-FEB-2000; 2000US-0180628P.			
PR	24-FEB-2000; 2000US-0184664P.			
PR	02-MAR-2000; 2000US-0186350P.			
PR	16-MAR-2000; 2000US-0189874P.			
PR	17-MAR-2000; 2000US-0190076P.			
PR	18-MAR-2000; 2000US-0198123P.			
PR	19-MAY-2000; 2000US-0205515P.			
PR	07-JUN-2000; 2000US-0209467P.			
PR	28-JUN-2000; 2000US-0214886P.			
PR	30-JUN-2000; 2000US-0215135P.			
PR	07-JUL-2000; 2000US-0216647P.			
XX				



CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention

SQ Sequence 25576 BP; 7153 A; 6020 C; 6060 G; 6343 T; 0 U; 0 Other;  
Query Match 83.2%; Score 20.8; DB 4; Length 25576;  
Best Local Similarity 91.7%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACAGCCGAGCATG 24

Db 12225 AAAAAAAAAACAGCCGAGCGTG 12248

## RESULT 17

AAK83758

ID AAK83758 standard; DNA; 25576 BP.

XX AAK83758;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38570.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 05-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.

PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 05-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-483426/52.  
 DR  
 XX  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 PT  
 XX Disclosure; SEQ ID NO 38570; 3071pp + Sequence Listing; English.  
 PS  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention  
 CC  
 XX Sequence 25576 BP; 7153 A; 6020 C; 6060 G; 6343 T; 0 U; 0 Other;  
 SQ  
 Query Match 83.2%; Score 20.8; DB 4; Length 25576;  
 Best Local Similarity 91.7%; Pred. No. 44;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCAGGCCAGGCGATG 24  
 Db 12225 AAAAAAAAAATCAGGCCAGGCGATG 12248  
 RESULT 18  
 AAL05618

ID AAL05618 standard; DNA; 25576 BP.  
 XX  
 AC AAL05618;  
 XX  
 DT 21-NOV-2001 (first entry)  
 XX  
 DE Human reproductive system related antigen DNA SEQ ID NO: 8306.  
 XX Human reproductive system related antigen; reproductive system disorder;  
 KW Human; cancer; gene therapy; ds.  
 KW Homo sapiens.  
 OS  
 XX  
 PN WO200155320-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001339.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214866P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0233196P.  
 PR 14-SEP-2000; 2000US-0233297P.  
 PR 14-SEP-2000; 2000US-0233298P.  
 PR

PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 23-SEP-2000; 2000US-0234997P.  
PR 23-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236376P.  
PR 29-SEP-2000; 2000US-0236388P.  
PR 29-SEP-2000; 2000US-0236399P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
PT  
XX  
XX Disclosure; SEQ ID NO 8306; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
SQ Sequence 25576 BP; 7153 A; 6020 C; 6060 G; 6343 T; 0 U; 0 Other;  
  
Query Match 83.2%; Score 20.8; DB 4; Length 25576;  
Best Local Similarity 91.7%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAATCAGCGCCAGGCATG 24  
Db 12225 AAAAAAAAAATCAGCGCCAGGCATG 12248  
  
RESULT 19  
ACN44934\_1/c  
Continuation (2 of 4) of ACN44934 from base 100001 (Human genomic sequence hCG1811271.  
WP Sequence split into 4 fragments LOCUS ACN44934 Accession ACN44934  
WP Fragment Name Begin End  
WP ACN44934\_0 1 110000  
WP ACN44934\_1 100001 210000  
WP ACN44934\_2 200001 310000  
WP ACN44934\_3 300001 401616  
  
Query Match 83.2%; Score 20.8; DB 11; Length 110000;  
Best Local Similarity 91.7%; Pred. No. 57;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 AAAAAAAAAATCAGCGCCAGGCATG 25  
Db 40060 AAAAAAAAAATCAGCGCCAGGCATG 40037  
  
RESULT 20  
ADZ46976\_3/c  
Continuation (4 of 6) of ADZ46976 from base 300001 (Human phosphodiesterase 4B DNA SEQ  
WP Sequence split into 6 fragments LOCUS ADZ46976 Accession ADZ46976  
WP Fragment Name Begin End  
WP ADZ46976\_0 1 110000  
WP ADZ46976\_1 100001 210000  
WP ADZ46976\_2 200001 310000  
WP ADZ46976\_3 300001 410000  
WP ADZ46976\_4 400001 510000  
WP ADZ46976\_5 500001 599001  
  
Query Match 83.2%; Score 20.8; DB 14; Length 110000;  
Best Local Similarity 91.7%; Pred. No. 57;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;





PF 21-FEB-2000; 2000EP-00200610.  
 XX  
 PR  
 XX 26-FEB-1999; 99US-0122487P.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 XX WPI; 2000-500381/45.  
 DR  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 PT  
 XX Claim 1; SEQ ID NO 18230; 71pp + Sequence Listing; English.  
 PS  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors  
 CC  
 XX SQ Sequence 206 BP; 47 A; 36 C; 42 G; 76 T; 0 U; 5 Other;  
 Query Match 80.8%; Score 20.2; DB 3; Length 206;  
 Best Local Similarity 88.0%; Pred. No. 36;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCAGGCCAGCATGG 25  
 Db 201 AAAAAAAAAATCAGGCCAGCATGG 177  
 RESULT 24  
 ACL56670/c  
 ID ACL56670 standard; cDNA; 302 BP.  
 XX  
 AC ACL56670;  
 XX  
 XX 24-MAR-2005 (first entry)  
 DT  
 XX Human colon cancer differentially expressed polynucleotide, SEQ ID:2805.  
 DE  
 XX Differential expression; diagnosis; therapy; drug screening; cancer;  
 XX neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;  
 KW ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO2005000087-A2.  
 PN  
 XX  
 XX 06-JAN-2005.  
 PD  
 XX  
 XX 13-MAY-2004; 2004WO-US015421.  
 PF  
 XX  
 XX 03-JUN-2003; 2003US-0475872P.  
 PR  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Randazzo F, Moler E, Escobedo J, Garcia PD;  
 PI  
 XX WPI; 2005-075421/08.  
 DR  
 XX  
 XX New isolated polynucleotides, which are differentially expressed in colon

cancer cell, useful for treating cancer, e.g. colon cancer, breast cancer, or pancreatic cancer.  
 Claim 1; SEQ ID NO 2805; 97pp; English.  
 The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which are differentially expressed in colon cancer cells. The invention also relates to vectors and host cells comprising a differentially expressed polynucleotide of the invention; a method for detecting a cancerous cell by detection of a gene product of the polynucleotides; a method for inhibiting a cancerous phenotype of a cell by inhibiting a gene product of the polynucleotides; a method of treating an individual with cancer by administration of a modulator of a gene product of the polynucleotides; and an isolated antibody that specifically binds to a polypeptide encoded by one of the 9672 polynucleotides. The polynucleotides, polypeptides, antibodies, and methods are useful for the detection of cancerous cells; for the diagnosis, prognosis and management of cancer; for the identification of agents that modulate the phenotype of cancerous cells; for the identification of therapeutic targets for cancer chemotherapy; and for the treatment of cancer, especially colon cancer and metastasized colon cancer, but also breast or pancreatic cancer. The polynucleotides are also useful as a source of probes or primers for use in diagnostic methods. The differentially expressed polynucleotides or their encoded proteins can additionally be used as vaccines to modulate primary immune responses for the prevention or treatment of cancer. The present sequence represents a specifically claimed polynucleotide which is differentially expressed in colon cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 302 BP; 76 A; 61 C; 61 G; 104 T; 0 U; 0 Other;  
 Query Match 80.8%; Score 20.2; DB 14; Length 302;  
 Best Local Similarity 88.0%; Pred. No. 39;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCAGGCCAGCATGG 25  
 Db 139 AAAAAAAAAAGGCCAGCATGG 115  
 RESULT 25  
 AAK74375  
 ID AAK74375 standard; DNA; 339 BP.  
 XX  
 AC AAK74375;  
 XX  
 XX 07-NOV-2001 (first entry)  
 DT  
 XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:29187.  
 DE  
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200157182-A2.  
 PN  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX  
 XX 17-JAN-2001; 2001WO-US001354.  
 PF  
 XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0231124P.  
PR 08-SEP-2000; 2000US-0231143P.  
PR 08-SEP-2000; 2000US-0231144P.  
PR 08-SEP-2000; 2000US-0231145P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231568P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-023637P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 29187; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to



CC recombination. the protein can be used to identify other proteins  
CC involved in development and progression of macular degeneration. The  
CC present sequence is a genomic DNA fragment containing an exon of a retina  
CC specific gene of the invention

XX  
SQ Sequence 384 BP; 95 A; 85 C; 86 G; 118 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 384;

Best Local Similarity 88.0%; Pred. No. 40;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 125 AAGAAATTCGCGCCAGGCATGG 101

RESULT 28

AAK83882/C

ID AAK83882 standard; DNA; 392 BP.

XX AC AAK83882;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38694.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001354.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 23-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.

PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.

PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.

PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.

PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.

PR 17-NOV-2000; 2000US-0249208P.

PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.

```
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254057P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 38694; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 392 BP; 100 A; 79 C; 79 G; 134 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 392;
XX Best Local Similarity 88.0%; Pred. No. 40;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
XX ||||| ||||| ||||| |||||
XX Db 208 AAAAAAAAAAAAAAGGCCAGGCATGG 184
XX
XX RESULT 29
XX AAI87906/c
XX ID AAI87906 standard; cDNA; 403 BP.
XX
XX AC AAI87906;
XX
```

---

```
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 7966.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO07975.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 7966; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO0010-AAO3910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 403 BP; 86 A; 120 C; 95 G; 102 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 403;
XX Best Local Similarity 88.0%; Pred. No. 41;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
XX ||||| ||||| ||||| |||||
XX Db 94 AAAAAAAAAAAAAAGGCCAGGCATGG 70
XX
XX RESULT 30
XX ABL82000/c
XX ID ABL82000 standard; cDNA; 465 BP.
XX
XX AC ABL82000;
XX
XX 17-MAY-2002 (first entry)
XX
XX Human ovarian cancer related cDNA clone SEQ ID NO:4978.
XX
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192581-A2.
XX
XX
```



PA (HYSE-) HYSEQ INC.  
XX Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I,  
XX WPI; 2002-241905/29.  
XX  
XX New nucleic acid for producing a polypeptide, detecting differentially  
PT expressed genes correlated with a cancerous state of a mammalian cell,  
PT and inhibiting tumor growth.  
XX  
XX Claim 1; SEQ ID NO 1717; 883pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
CC with cytostatic activity. The polynucleotide is used to produce a  
CC polypeptide, to detect differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell and to inhibit tumour growth. The  
CC polynucleotide is used as a probe in mapping and tissue profiling. The  
CC encoded polypeptide and antibodies to the polypeptide can also be used  
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
CC gene therapy. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 480 BP; 158 A; 97 C; 132 G; 93 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 20.2; DB 6; Length 480;  
Best Local Similarity 88.0%; Pred. No. 42;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCCGCGCATGG 25  
Db 238 AAAAAAAAAATCAGCGCCGCGCATGG 262  
RESULT 33  
ACL60088  
ID ACL60088 standard; CDNA; 480 BP.  
XX  
XX ACL60088;  
XX  
XX 24-MAR-2005 (first entry)  
XX  
XX Human colon cancer differentially expressed polynucleotide, SEQ ID:6223.  
XX  
XX Differential expression; diagnosis; therapy; drug screening; cancer;  
KW neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;  
KW ss.  
XX Homo sapiens.  
XX  
XX WO2005000087-A2.  
XX  
XX 06-JAN-2005.  
XX  
XX 13-MAY-2004; 2004WO-US015421.  
XX  
XX 03-JUN-2003; 2003US-0475872P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Randazzo F, Moler E, Escobedo J, Garcia PD;  
PI  
XX WPI; 2005-075421/08.  
XX  
XX New isolated polynucleotides, which are differentially expressed in colon  
PT cancer cell, useful for treating cancer, e.g. colon cancer, breast  
PT cancer, or pancreatic cancer.  
XX  
XX Claim 1; SEQ ID NO 6223; 97pp; English.  
XX  
XX The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which  
CC are differentially expressed in colon cancer cells. The invention also

CC relates to vectors and host cells comprising a differentially expressed  
CC polynucleotide of the invention; a method for detecting a cancerous cell  
CC by detection of a gene product of the polynucleotides; a method for  
CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product  
CC of the polynucleotides; a method of treating an individual with cancer by  
CC administration of a modulator of a gene product of the polynucleotides;  
CC and an isolated antibody that specifically binds to a polypeptide encoded  
CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,  
CC antibodies, and methods are useful for the detection of cancerous cells;  
CC for the diagnosis, prognosis and management of cancer; for the  
CC identification of agents that modulate the phenotype of cancerous cells;  
CC for the identification of therapeutic targets for cancer chemotherapy;  
CC and for the treatment of cancer, especially colon cancer and metastasized  
CC colon cancer, but also breast or pancreatic cancer. The polynucleotides  
CC are also useful as a source of probes or primers for use in diagnostic  
CC methods. The differentially expressed polynucleotides or their encoded  
CC proteins can additionally be used as vaccines to modulate primary immune  
CC responses for the prevention or treatment of cancer. The present sequence  
CC represents a specifically claimed polynucleotide which is differentially  
CC expressed in colon cancer. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 480 BP; 158 A; 97 C; 132 G; 93 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 20.2; DB 14; Length 480;  
Best Local Similarity 88.0%; Pred. No. 42;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCCGCGCATGG 25  
Db 238 AAAAAAAAAATCAGCGCCGCGCATGG 262  
RESULT 34  
ABA59354  
ID ABA59354 standard; DNA; 484 BP.  
XX  
XX ABA59354;  
XX  
XX 01-FEB-2002 (first entry)  
XX  
XX Human foetal liver single exon nucleic acid probe #7659.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human fetal liver.  
XX  
XX Claim 1; SEQ ID NO 7659; 639pp + Sequence Listing; English.  
PS



xx The invention relates to a single exon nucleic acid probe for measuring  
cc human gene expression in a sample derived from human foetal liver. The  
cc single exon nucleic acid probes may be used for predicting, measuring and  
cc displaying gene expression in samples derived from human fetal liver. The  
cc present sequence is a single exon nucleic acid probe of the invention.  
cc Note: The sequence data for this patent did not form part of the printed  
cc specification, but was obtained in electronic format directly from WIPO  
cc at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
xx  
xx Sequence 484 BP; 166 A; 95 C; 138 G; 85 T; 0 U; 0 Other;

Query Match	80.8%;	Score 20.2;	DB 4;	Length 484;
Best Local Similarity	88.0%;	Pred. No. 42;		
Matches 22:	Conservative	0;	Mismatches 3;	Indels 0; Gaps 0

**Qy**

1 AAAAAAAAAATCAGGCCAGGCATGG 25  
||||| ||||| ||||| ||||| |||||

**pB**

55 AAAAAACAAGCAAGGCCAGGCATGG 79

RESULT	35
AAI39155	
ID	AAI39155 standard; DNA; 484 BP.
XX	
XX	
AC	AAI39155;
XX	
XX	
DT	17-OCT-2001 (first entry)
XX	
DE	Probe #7841 used to measure gene expression in human placenta sample.
XX	
KW	Probe; microarray; human; placenta; antenatal diagnosis;
KW	genetic disorder; ss.
KW	

Claim 25; SEQ ID NO 7841; 654pp; English.

Query Match	80.8%	Score 20.2;	DB 4;	Length 484;
Best Local Similarity	88.0%;	Pred. No. 42;		
Matches 22: Conservative	0: Mismatches	3: Indels	0: Gaps	0:

Qy 1 AAAAAAAAAATCACGGCCAGGCATGG 25  
pB 55 AAAAAACAAGCAAGGCCAGGCATGG 79

RESULT 36  
AAK33368  
ID AAK33368 standard; DNA; 484 BP.  
XX  
AC AAK33368;

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
KW

XX  
OS  
Homo sapiens.

XX PN WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000668.

AA  
PR 04-FEB-2000; 2000US-0180312P.

PR Z6-MAY-2000; 2000US-0207430F;  
PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00832388;  
PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236333F;  
PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS IN

XX Penn SG, Hanzel DK, Chen W,

XX  
DR WPI: 2001-488900/53.

XX  
PS Example 4: SEQ ID NO 7925: 658bp + Sequence Listing: English.

xx The present invention provides a number of single exon nucleic acid  
cc probes which are derived from genomic sequences expressed in the human  
cc bone marrow. They can be used to measure gene expression in bone marrow  
cc samples, which may enable the improved diagnosis and treatment of cancers  
cc such as lymphoma, leukaemia and myeloma. The present sequence is one of  
cc the probes of the invention

```
Query Match      80.8%; Score 20.2; DB 4; Length 484;
Best Local Similarity 88.0%; Pred. No. 42;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
||| ||| ||| ||| ||| ||| ||| |||  
Db 55 AAAAAACAAGCAAGGCCAGGCATGG 79

RESULT 37  
AAK07570  
ID AAK07570 standard: DNA: 484 BP.

AA  
AC AAK07570;

DT 05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 7561.

KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
KW ss.

XX Homo sapiens.

OS WO200157275-A2.

PN WO200157275-A2.

PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.

PT Example 4; SEQ ID NO 7561; 650pp + Sequence Listing; English.  
PS The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention

XX Sequence 484 BP; 166 A; 95 C; 138 G; 85 T; 0 U; 0 Other;

XX Query Match 80.8%; Score 20.2; DB 4; Length 484;

XX Best Local Similarity 88.0%; Pred. No. 42;

XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

DB 55 AAAAAACAAGCAGCGCCAGGCATGG 79

RESULT 38

ABS33124

ID ABS33124 standard; DNA; 484 BP.

XX ABS33124;

AC ABS33124;

XX 25-FEB-2003 (first entry)

DT Human liver single exon probe, SEQ ID No 8114.

DE Human; single exon nucleic acid probe; liver; cirrhosis;

XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;

XX coronary heart disease; ss.

XX Homo sapiens.

OS WO200157273-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.

PR

PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.

XX Claim 1; SEQ ID NO 8114; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/ fragments). The probe hybridises at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABS25011-ABS51005 represent human  
XX liver single exon nucleic acid probes of the invention. Note: The  
XX sequence information for this patent does not appear in the printed  
XX specification but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

XX Query Match 80.8%; Score 20.2; DB 4; Length 484;

XX Best Local Similarity 88.0%; Pred. No. 42;

XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

DB 55 AAAAAACAAGCAGCGCCAGGCATGG 79

RESULT 39

ABS08206

ID ABS08206 standard; DNA; 484 BP.

XX ABS08206;

XX 19-AUG-2002 (first entry)

DT Human genome-derived single exon probe from lung SEQ ID No 8197.

DE Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

XX chronic obstructive pulmonary disease; interstitial lung disease;

XX familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;

XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX primary ciliary dyskinesia; fibrocystic pulmonary dysplasia;

XX hyaline membrane disease.

XX Homo sapiens.

OS WO200186003-A2.

PN 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

PR



OS Homo sapiens.  
XX WO200009525-A2.  
XX  
XX  
PD 24-FEB-2000.  
XX  
XX 03-AUG-1999; 99WO-US017712.  
XX  
XX 03-AUG-1998; 98US-0095212P.  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA  
XX  
XX Nyce JW;  
XX WPI; 2000-205971/18.  
XX  
XX New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers.  
XX  
XX Disclosure; Page 1075-1076; 1343pp; English.  
XX  
XX The present invention describes a new composition comprising an antisense  
CC oligonucleotide (ON) with low adenine (up to 15%), which targets  
CC nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have antiinflammatory, antiallergic,  
CC antiasthmatic, cytostatic and analgesic activities. The compositions are  
CC useful for the treatment of diseases associated with inflammation,  
CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,  
CC impaired respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
CC carcinomas, and cancers which may metastasise to the lungs, including  
CC breast and prostate cancer. The reduction of the adenine content of the  
CC ONs reduces side effects. The A-containing ONs break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing  
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
CC nucleotide sequences given in the sequence listing from the present  
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185  
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ  
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to  
CC AAA33922) are specifically claimed ONs from the present invention. N.B.  
CC Sequences given in the disclosure of the present invention do not match  
CC up with their corresponding SEQ ID NO: sequences given in the sequence  
CC listing  
XX  
XX Sequence 666 BP; 206 A; 165 C; 174 G; 121 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 20.2; DB 3; Length 666;  
Best Local Similarity 88.0%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCCAGGCATGG 25  
Db 99 AAAAAAAAAAAAGCCAGGCATGG 123  
RESULT 42  
AAF21242  
ID AAF21242 standard; DNA; 666 BP.  
XX  
AC AAF21242;  
XX  
XX 14-MAR-2001 (first entry)  
XX  
XX Human low adenine antisense oligonucleotide related sequence #2809.  
DE  
XX Low adenine antisense oligonucleotide; phosphorothioate; allergy;  
XX human; airway disorder; bronchoconstriction; lung inflammation;  
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;

KW immunosuppressive; antiaesthmatic; analgesic; hypotensive; cytostatic;  
KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;  
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2000062736-A2.  
PN  
XX  
XX 26-OCT-2000.  
PD  
XX  
XX 24-MAR-2000; 2000WO-US008020.  
PF  
XX  
XX 06-APR-1999; 99US-0127958P.  
PR  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA  
XX  
XX (NYCE/) NYCE J W.  
XX  
XX Nyce JW;  
PI  
XX  
XX WPI; 2000-679539/66.  
DR  
XX  
XX Low adenine (A) content antisense oligonucleotides which do not trigger  
PT adenosine receptors during metabolism, useful e.g. for treating cancers  
PT and respiratory obstructions.  
PT  
XX  
XX Disclosure; Page 1155; 1592pp; English.  
XX  
XX The present invention describes low adenine (A) content antisense  
CC oligonucleotides and compositions (I) comprising them. In the antisense  
CC oligonucleotides the A is replaced by a 'universal' or alternative base.  
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
CC immunosuppressive, antiaesthmatic, hypotensive and cytostatic activities.  
CC The antisense oligonucleotides and (I) can be used to down-regulate the  
CC expression and/or activity of target polypeptides associated with  
CC lung/respiratory disorders and malignancies, such as stimulating and  
CC activating peptide factors and transmitters, transcription factors,  
CC immunoglobulins and antibodies, antibody receptors, cytokines and  
CC chemokines, endogenously produced specific and non-specific enzymes,  
CC binding proteins, adhesion molecules and their receptors, cytokine and  
CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
CC nervous system (CNS) and peripheral nervous and non-nervous system  
CC receptors, CNS and peripheral nervous and non-nervous system peptide  
CC transmitters, defensins, growth factors, vasoactive peptides and  
CC receptors, binding proteins and malignancy associated proteins. The  
CC antisense oligonucleotides may be used in this way to treat disorders  
CC including respiratory obstruction (especially pulmonary obstruction  
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or  
CC surfactant hypoproduction which are associated with a disease or  
CC condition selected from pulmonary vasoconstriction, inflammation,  
CC allergies, asthma, impeded respiration, respiratory distress syndrome  
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
CC fragments and antisense oligonucleotides used in the exemplification of  
CC the present invention  
XX  
XX Sequence 666 BP; 206 A; 165 C; 174 G; 121 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 20.2; DB 3; Length 666;  
Best Local Similarity 88.0%; Pred. No. 44;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCCAGGCATGG 25  
Db 99 AAAAAAAAAAAAGCCAGGCATGG 123  
RESULT 43

```
ABZ96936
ID ABZ96936 standard; DNA; 666 BP.
XX
AC ABZ96936;
XX
XX 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.
XX
PA (EPITG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12178; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antinflammatory steroid and ubiquinone. A composition of the invention
CC has antinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 666 BP; 206 A; 165 C; 174 G; 121 T; 0 U; 0 Other;
XX
Query Match 80.8%; Score 20.2; DB 10; Length 666;
Best Local Similarity 88.0%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25
| | | | | | | | | | | | | | | | | | | | |
Db 99 AAAAAAAAAAAAAAGGCGCAGCATGG 123
| | | | | | | | | | | | | | | | | | | | |
RESULT 44
```

```
SQ Sequence 666 BP; 206 A; 165 C; 174 G; 121 T; 0 U; 0 Other;
Query Match      80.8%; Score 20.2; DB 11; Length 666;
Best Local Similarity 88.0%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   ||||| ||| ||||| |||||
DB 99 AAAAAAAAAAAGGCCAGGCATGG 123

RESULT 45
AAS28505/c
ID AAS28505 standard; DNA; 1131 BP.
XX
AC AAS28505;
XX
DT 07-NOV-2001 (first entry)
XX
DE Genomic sequence #345 encoding for novel human respiratory antigen.
XX
KW Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; oilyfactory;
KW respiratory active; ds.
XX
OS Homo sapiens.
XX
PN WO20015448-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001333.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216880P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217497P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-023937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
```

PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-476224/51.  
XX  
XX Isolated polypeptide for treating, preventing and/ or prognosing  
PT disorders related to the respiratory system including respiratory cancers  
PT and also for testing and detection e.g. diagnosis.  
XX  
XX Disclosure; SED ID No 939; 546pp; English.  
XX  
XX The present invention relates to the isolation of novel human respiratory  
CC antigens (AAU17685-AAU17975), and cDNA and genomic sequences encoding for  
CC these polypeptides. The sequences of the invention are useful for  
CC preventing, treating and/or prognosing disorders related to the  
CC respiratory system including throat disorders (e.g. vocal cord paralysis,  
CC tonsillitis, and laryngitis), lung disorders e.g. pneumonia, allergic  
CC disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose  
CC disorders and cancers of the respiratory tissues e.g. lung cancer. The  
CC polynucleotide sequences of the invention are useful in gene therapy and  
CC antisense therapy. AAS28161-AA328764 represent genomic sequences encoding  
CC for novel human respiratory antigens. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1131 BP; 311 A; 240 C; 271 G; 309 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 20.2; DB 4; Length 1131;  
Best Local Similarity 88.0%; Pred. No. 48;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 500 AAAAAAAAAATTAAGGCCAGGCATGG 476  
RESULT 46  
AAS29996/c  
ID AAS29996 standard; DNA; 1131 BP.  
XX  
XX AAS29996;  
XX  
XX 21-NOV-2001 (first entry)  
DT  
XX Human lung antigen genomic DNA #66.  
DE  
XX

KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW hyperphalmmological; vulnery; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; food preservative;  
KW tissue regeneration; anti-infertility; food additive.  
XX Homo sapiens.  
XX WO200155303-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001301.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.



```
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 23-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.

PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX respiratory disorders related to the lung including lung cancers and also
XX for testing and detection e.g. diagnosis.
XX
XX Claim 1; SEQ ID NO 260; 507pp; English.
XX
XX Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
XX the lung antigen polypeptides of the invention. Lung antigen polypeptides
XX and their associated polynucleotides are useful in the diagnosis,
XX treatment and prevention of various types of disorders in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
XX pathological condition can be determined by detecting the presence or
XX absence of a mutation in a lung antigen polynucleotide. The treatable
XX disorders include autoimmune diseases such as rheumatoid arthritis,
XX hyperproliferative disorders such as neoplasms of the breast or liver,
XX cardiovascular disorders such as cardiac arrest, cerebrovascular
XX disorders such as cerebral ischaemia, nervous system disorders such as
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi,
XX ocular disorders such as corneal infection, endocrine disorders such as
XX premature labour and infertility, gastrointestinal disorders such as
XX Crohn's disease, renal disorders such as asthma and pleurisy. The polypeptides can
XX also be used to aid wound healing, to prevent skin aging due to sunburn,
XX to maintain organs before transplantation, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

Query Match 80.8%; Score 20.2; DB 5; Length 1131;
Best Local Similarity 88.0%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGCGCATGG 25
Db 500 AAAAAAAAAATTAAGGCCAGCGCATGG 476

RESULT 47
ADB33333/c
ID ADB33333 standard; DNA; 1131 BP.
XX
XX ADB33333;
XX
XX 04-DEC-2003 (first entry)
XX
XX Human novel lung related polypeptide DNA SEQ ID NO 260.
XX
XX gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
XX adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
XX immunodeficiency; X-linked agammaglobulinaemia;
XX X-linked infantile agammaglobulinaemia; inflammatory disorder;
XX adrenalitis; alveolitis; immune complex disease; serum sickness;
XX polyarteritis nodosa; bleeding disorder; thrombocytopenia;
XX Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
```

KW multiple myeloma; macrophage related disorder; Gaucher's disease;  
KW Neimann-Pick disease; tumour; colon cancer; pancreatic cancer;  
KW renal disorder; nephritis; bone disorder; Albers-Schönberg disease;  
KW bowleg; muscle disorder; Becker's muscular dystrophy;  
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;  
KW traumatic lesion; endocrine disorder; Cushing's syndrome;  
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;  
KW gastric reflux; human; ds.

XX Homo sapiens.

XX US2003054368-A1.

XX 20-MAR-2003.

XX 22-FEB-2002; 2002US-00079854.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216680P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-023637P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0244675P.  
PR 08-NOV-2000; 2000US-0244676P.  
PR 08-NOV-2000; 2000US-0244677P.  
PR 08-NOV-2000; 2000US-0244678P.  
PR 08-NOV-2000; 2000US-0245213P.  
PR 08-NOV-2000; 2000US-024524P.  
PR 08-NOV-2000; 2000US-024525P.  
PR 08-NOV-2000; 2000US-024526P.  
PR 08-NOV-2000; 2000US-024527P.  
PR 08-NOV-2000; 2000US-024528P.  
PR 08-NOV-2000; 2000US-024532P.  
PR 08-NOV-2000; 2000US-0246603P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.

```
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764878.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-695900/66.
XX
XX Novel isolated lung antigen polypeptides useful for treating, preventing,
PT diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
PT Von Willebrand's disease.
XX
XX Disclosure; SEQ ID NO 260; 178pp; English.
XX
XX The invention relates to an isolated lung antigen polypeptide sequence or
CC encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
CC are useful for treating, preventing, diagnosing and/or prognosing
CC diseases and/or disorders such as pathological cell proliferative
CC neoplasias e.g. acute myelogenous leukaemias, adenocarcinoma; respiratory
CC disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
CC X-linked agammaglobulinaemia, X-linked infantile agammaglobulinaemia;
CC inflammatory disorders such as adrenailitis, alveolitis; immune complex
CC diseases such as serum sickness, polyarteritis nodosa; bleeding disorders
CC such as thrombocytopenia, Von Willebrand's disease; acquired platelet
CC dysfunction such as kidney failure, multiple myeloma; disorders
CC associated with macrophage numbers and/or macrophage function such as
CC Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
CC pancreatic cancer; renal disorders such as kidney failure, nephritis;
CC bone disorders such as Albers-Schonberg disease, bowlegs; muscle
CC disorders such as Becker's muscular dystrophy, Duchenne's muscular
CC dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
CC ; endocrine disorders such as Cushing's syndrome, corticosteroid

Query Match 80.8%; Score 20.2; DB 10; Length 1131;
Best Local Similarity 88.0%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCAGGCATCG 25
| | | | | | | | | | | | | | | | | |
Db 500 AAAAAAAAAATTAAGGCGCAGGCATCG 476

RESULT 48
ADG41701/c
ID ADG41701 standard; DNA; 1131 BP.
XX
AC ADG41701;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human respiratory system associated genomic DNA seq id 939.
XX
XX antiinflammatory; antiallergic; antiasthmatic; cytostatic; gene therapy;
KW human respiratory system antigen;
KW human respiratory system associated polynucleotide;
KW respiratory system disorder; throat disorder; vocal cord paralysis;
KW tonsillitis; laryngitis; lung disorder; pneumonia; allergic disorder;
KW asthma; eosinophilic pneumonia; pleurisy; cystic fibrosis; emphysema;
KW histiocytosis; sarcoidosis; nose disorder; rhinitis; sinusitis; neoplasm;
KW cancer; respiratory tissue cancer; throat cancer; lung cancer;
KW cancer of the nose; gene therapy; chromosome identification; forensic;
KW human respiratory system associated protein; ds; human.
XX
OS Homo sapiens.
XX
XX US2003215893-A1.
FN
```

```
XX 20-NOV-2003.
PD 07-AUG-2002; 2002US-00212872.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 28-JUN-2000; 2000US-0209467P.
PR 30-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216447P.
PR 11-JUL-2000; 2000US-0216880P.
PR 14-JUL-2000; 2000US-0217487P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226968P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229145P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 28-SEP-2000; 2000US-0235935P.
PR
```

```
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764860.
PR 14-FEB-2002; 2002US-00074095.
(PUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-902033/82.
XX Novel respiratory system antigen and polynucleotides encoding the
PT polypeptides, useful for treating diagnosing, treating or preventing
PT tonsillitis, pneumonia, asthma and cystic fibrosis, emphysema, throat
PT cancer.
XX Disclosure; SEQ ID NO 939; 236pp; English.
XX The invention describes an isolated polypeptide (I) comprising an amino
CC acid sequence that is at least 90% identical to polypeptide fragment of
CC any one of 299 respiratory system antigen sequences (PS) and having
CC biological activity, polypeptide domain or epitope of PS, full-length
CC protein of PS, or variant, allelic variant or species homolog of PS. (I)
CC or a polynucleotide (II) encoding (I) is also useful for diagnosing a
CC pathological condition or a susceptibility to a pathological condition in
CC a subject which involves determining the presence or absence of mutation
CC in (II) or determining the presence or amount of expression of (I) in a
CC biological sample and diagnosing a pathological condition based on the
CC result. The human respiratory system associated polynucleotides, the
CC polypeptides encoded by them, and antibodies that immunospecifically bind
CC these polypeptides are useful in diagnosis, treatment, prevention and/or
CC prognosis of disorders of respiratory system such as throat disorders
CC (e.g., vocal cord paralysis, tonsillitis, and laryngitis), lung disorders
CC (e.g., pneumonia), allergic disorders, (e.g., asthma and eosinophilic
CC pneumonia), pleurisy, cystic fibrosis, emphysema, histiocytosis,
CC sarcoidosis, nose disorders (rhinitis and sinusitis), neoplasms and/or
CC cancers of respiratory tissues (e.g., throat cancer, lung cancer, and
CC cancer of the nose). The polynucleotides are useful in gene therapy
CC techniques, for chromosome identification, identifying individuals from
CC
Query Match 80.8%; Score 20.2; DB 10; Length 1131;
Best Local Similarity 88.0%; Pred. No. 49;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 500 AAAAAAAAAATTAAGGCCAGGCATGG 476
RESULT 49
AD197475/C
ID AD197475 standard; DNA; 1131 BP.
XX
AC AD197475;
XX
DT 04-NOV-2004 (first entry)
XX
DE Human respiratory system associated polypeptide-related DNA SeqID939.
XX
KW respiratory system-related polypeptide; antiasthmatic; antibacterial;
KW antiinflammatory; cytostatic; antianaemic; antiallergic; gene therapy;
KW pneumonia; lung cancer; cystic fibrosis; asthma; sarcoidosis; rhinitis;
KW anaemia; leukaemia; inflammation; sinusitis;
KW Chronic obstructive pulmonary disease; infectious disease; human; ds.
XX
OS Homo sapiens.
XX
PN US2003077704-A1.
XX
PD 24-APR-2003.
XX
PF 14-FEB-2002; 2002US-00074095.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
```

PR 19-MAY-2000; 2000US-020551SP.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215113P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232337P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234987P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 02-OCT-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764860.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-765403/72.  
XX  
XX New human respiratory system-related polypeptide and genes, useful for  
XX treating, preventing or diagnosing e.g. pneumonia, lung cancer, cystic  
XX fibrosis, asthma, sarcoidosis, rhinitis, leukemia, inflammations or  
XX sinusitis.  
XX  
XX Disclosure; SEQ ID NO 939; 202pp; English.  
XX  
XX This invention is related to a novel isolated polypeptide, which  
XX comprises a human respiratory system-related polypeptide, and the DNA  
CC



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-1

Perfect score: 25

Sequence: 1 aaaaaaaatcacggccagcatgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_est9.\*

10: gb\_est10.\*

11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	246	1	BB056745
2	25	100.0	329	1	AW413970
3	25	100.0	410	1	AW763628
4	25	100.0	469	7	CK343886
5	25	100.0	534	2	BB693815
6	25	100.0	649	6	CB248873
7	23.4	93.6	354	8	T69797
8	23.4	93.6	389	1	AA507708
9	22.4	89.6	413	9	A2008849
10	21.8	87.2	254	1	AW554415
11	21.8	87.2	317	2	BP469089
12	21.8	87.2	322	1	BB105780
13	21.8	87.2	330	7	CB671031
14	21.8	87.2	372	5	BY690105
15	21.8	87.2	373	1	AW492561
16	21.8	87.2	375	2	BB791462
17	21.8	87.2	379	5	BY698909
18	21.8	87.2	386	7	CN698559
19	21.8	87.2	388	5	BY425365
20	21.8	87.2	392	5	BY684912
21	21.8	87.2	394	5	BY691014
22	21.8	87.2	401	5	BY701163

C 23	21.8	87.2	403	5	BY694371
C 24	21.8	87.2	404	5	BY690774
C 25	21.8	87.2	415	5	BY655013
C 26	21.8	87.2	418	2	BG074809
C 27	21.8	87.2	418	8	W12097
C 28	21.8	87.2	421	1	AI124789
C 29	21.8	87.2	423	5	BY697032
C 30	21.8	87.2	425	2	BB781154
C 31	21.8	87.2	425	2	BB789804
C 32	21.8	87.2	426	5	BY383430
C 33	21.8	87.2	431	1	AA744879
C 34	21.8	87.2	431	1	AI642924
C 35	21.8	87.2	434	5	BY693822
C 36	21.8	87.2	436	5	BY384796
C 37	21.8	87.2	447	2	BB782484
C 38	21.8	87.2	470	2	BB819202
C 39	21.8	87.2	474	3	BM205156
C 40	21.8	87.2	491	9	AZ450659
C 41	21.8	87.2	514	3	BM203574
C 42	21.8	87.2	519	1	AA721682
C 43	21.8	87.2	555	2	BB753963
C 44	21.8	87.2	575	3	BP768879
C 45	21.8	87.2	593	7	CK345513
C 46	21.8	87.2	595	1	AW554397
C 47	21.8	87.2	600	3	BP769047
C 48	21.8	87.2	630	10	AG075591
C 49	21.8	87.2	666	1	AW976597
C 50	21.8	87.2	715	7	CK787496
C 51	21.8	87.2	718	6	CB320467
C 52	21.8	87.2	821	8	CK757930
C 53	21.8	87.2	962	5	BA461680
C 54	21.8	87.2	1030	1	AL579068
C 55	21.8	87.2	3341	4	CR749847
C 56	21.8	87.2	3355	4	AK052299
C 57	21	84.0	1013	1	AL582650
C 58	20.8	83.2	345	9	AQ067376
C 59	20.8	83.2	355	1	AI420666
C 60	20.8	83.2	396	9	AQ285592
C 61	20.8	83.2	448	3	BJ582558
C 62	20.8	83.2	525	9	AQ037355
C 63	20.8	83.2	675	10	AG165060
C 64	20.8	83.2	709	9	AQ487377
C 65	20.8	83.2	719	10	CZ301420
C 66	20.8	83.2	728	9	AQ056247
C 67	20.8	83.2	835	9	BZ216878
C 68	20.8	83.2	1068	10	CZ390223
C 69	20.4	81.6	364	2	BF812239
C 70	20.4	81.6	398	2	BF811957
C 71	20.4	81.6	401	2	BF812248
C 72	20.2	80.8	141	9	AZ104693
C 73	20.2	80.8	159	2	BG952210
C 74	20.2	80.8	155	6	CD172275
C 75	20.2	80.8	181	1	AA830764
C 76	20.2	80.8	210	1	AA669747
C 77	20.2	80.8	243	1	AU059999
C 78	20.2	80.8	264	7	CN277450
C 79	20.2	80.8	268	9	AQ267741
C 80	20.2	80.8	278	1	AW903860
C 81	20.2	80.8	278	2	BF903277
C 82	20.2	80.8	283	3	BM510815
C 83	20.2	80.8	286	1	AW833004
C 84	20.2	80.8	289	3	BI491315
C 85	20.2	80.8	289	1	AW020255
C 86	20.2	80.8	289	1	AW022442
C 87	20.2	80.8	289	1	AW022611
C 88	20.2	80.8	289	3	BI491894
C 89	20.2	80.8	289	3	BI492396
C 90	20.2	80.8	291	1	AW021440
C 91	20.2	80.8	294	7	CK300314
C 92	20.2	80.8	296	2	BB340383
C 93	20.2	80.8	298	2	BB488258
C 94	20.2	80.8	304	1	BB086006
C 95	20.2	80.8	315	1	BB098444

BY694371	BY694371
BY690774	BY690774
BY655013	BY655013
BG074809	H3139F01-
W12097	ma96f07.r1
AI124789	am56b10.x
BY697032	BY697032
BB781154	BB781154
BB789804	BB789804
BY383430	BY383430
AA744879	nv71h11.8
AI642924	vd62a08.x
BY693822	BY693822
BY384796	BY384796
BB782484	BB782484
BB819202	BB819202
BM205156	C0279F09-
AZ450659	1M024SD01
BM203574	C0256E04-
AA721682	ny72h09.8
BB753963	BB753963
BP768879	BP768879
CK345513	L0247D01-
AW554397	L0240805-
BP769047	BP769047
AG075591	Pan trogl
AW976597	EST388706
CK787496	AGENCOURT
CB320467	AGENCOURT
CK757930	AGENCOURT
BA461680	BA461680
AL579068	AL579068
CR749847	Homo sapi
AK052299	Mus muscu
AL582650	AL582650
AQ067376	HS-2237.A
AI420666	tf13c08.x
AQ285592	RPC111-89
BJ582558	BJ582558
AQ037355	CIT-HSP-2
AG165060	Pan trogl
AQ487377	RPC1-11-2
CZ301420	ZMMBF0076
AQ056247	CIT-HSP-2
BZ216878	CH230-467
CZ390223	ZMMBF0167
BF812239	QV1-C1017
BF811957	QV1-C1017
BF812248	QV1-C1017
AZ104693	RPCI-23-2
BG952210	CM4-CT062
CD172275	AGENCOURT
AA830764	OC53f07.8
AA669747	adale04.8
AU059999	AU059999
CN277450	170060001
AQ267741	RPCI11-69
AW903860	CM4-NM103
BF903277	CM2-MT019
BM510815	ij34e02.x
AW833004	RK3-CT000
BI491315	df07a08.w
AW020255	df07a08.y
AW022442	df38f12.y
AW022611	df41c11.y
BI491894	df16f02.y
BI492396	df23f07.w
AW021440	df23f07.y
CK300314	UI-E-EJ1-
BB340383	BB340383
BB488258	BB488258
BB086006	BB086006
BB098444	BB098444





```

Query Match      100.0%; Score 25; DB 1; Length 246;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
    |||||||
Db 100 AAAAAAAAAATCAGGCCAGGCATGG 124

RESULT 2
AW413970/c
LOCUS      AW413970
DEFINITION uc97dl1.x1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:2650485 3',
           mRNA sequence.
ACCESSION  AW413970
VERSION     AW413970.1 GI:6940118
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 329)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
           cDNA Library Preparation: Life Technologies, Inc.
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html

MGI:1030937
Seq primer: -40UP from Gibco
High quality sequence stop: 304.
Location/Qualifiers
FEATURES             source
   1..329
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="129,C57BL/6J,FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:2650485"
    /tissue_type="tumor, gross tissue"
    /dev_stage="10 months"
    /lab_hosts="DH10B"
    /clone_lib="NCI_CGAP Mam3"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
    Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
    Library constructed by Life Technologies. Investigators
    providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
    Reference for transgenic model: Xu et al., Nature Genetics
    22, 37-43 (1999)."
```

```

ORIGIN
Query Match      100.0%; Score 25; DB 1; Length 410;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
    |||||||
Db 151 AAAAAAAAAATCAGGCCAGGCATGG 127

RESULT 4
CKX43886/c
LOCUS      CKX43886
DEFINITION K0806E06-3 NIA Mouse 8.5-dpc whole Embryo CDNA Library (long) Mus
           musculus cDNA clone NIA:K0806E06 IMAGE:30078965 3', mRNA sequence.
ACCESSION  CKX43886
VERSION     CKX43886.1 GI:40299499
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 469)
AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE     Construction of long-transcript enriched cDNA libraries from
           submicrogram amounts of total RNAs by a universal PCR amplification
```

```

mRNA sequence.
AW763628
AW763628.1 GI:7695569
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 410)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: ur63e02.Y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

MGI:1057726
Seq primer: -40UP from Gibco
High quality sequence stop: 383.
Location/Qualifiers
FEATURES             source
   1..410
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="129,C57BL/6J,FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:3154970"
    /tissue_type="tumor, gross tissue"
    /dev_stage="10 months"
    /lab_hosts="DH10B"
    /clone_lib="NCI_CGAP Mam3"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
    Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
    Library constructed by Life Technologies. Investigators
    providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
    Reference for transgenic model: Xu et al., Nature Genetics
    22, 37-43 (1999)."
```



assistance we gratefully acknowledge."

ORIGIN	Query Match Best Local Similarity 100.0%; Score 25; DB 2; Length 534; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AAAAAAAAAATCAGGCCAGGCATGG 25       470 AAAAAAAAAATCAGGCCAGGCATGG 494 
Db	388 AAAAAAAAAATCAGGCCAGGCATGG 412 
RESULT 6	
LOCUS	CB248873
DEFINITION	UI-M-EXO-bvj-d-09-0-U1.r1 NIH_BMAP_EX0 Mus musculus cDNA clone IMAGE:5718872 5', mRNA sequence.
ACCESSION	CB248873
VERSION	CB248873.1 GI:28387712
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 649)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs-r@mail.nih.gov">cgabbs-r@mail.nih.gov</a> Tissue Procurement: Dr. James Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
FEATURES	source 1..649 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:5718872" /tissue_type="whole brain" /dev_stage="embryo 15.5 dpc" /lab_host="DH10B (T1 phage resistant)" /clone_lib="NIH_BMAP_EX0" /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is GTGGGTGGA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN	

```
Db      236 AAAAAAAAAATCGCCAGGCATGG 212

RESULT 8
AA507708/c
LOCUS
DEFINITION
ng85c03.s1 NCI CGAP Pr6 Homo sapiens cDNA clone IMAGE:941572
similar to contains_Alu repetitive element;contains MER2.t3 MER2
repetitive element ;, mRNA sequence.
AA507708
AA507708.1 GI:2244147
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 389)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.biol.llnl.gov/bbrp/image/image.html
Insert length: 632 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 356.

FEATURES
Location/Qualifiers
1..389
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:941572"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr6"
/notes="Vector: pAMP10; mRNA made from prostatic
intraepithelial neoplasia (low-grade), cDNA made by
oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
```

```
Query Match      93.6%; Score 23.4; DB 1; Length 389;
Best Local Similarity 96.0%; Pred. No. 24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCCAGGCATGG 25
|||||
Db 132 AAAAAAAAAATCGCCAGGCATGG 108

RESULT 9
AZ008849/c
LOCUS
DEFINITION
RPCI-23-375124-TV RPCI-23 Mus musculus genomic clone
RPCI-23-375124, genomic survey sequence.
AZ008849
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Query Match      93.6%; Score 22.4; DB 9; Length 413;
Best Local Similarity 95.8%; Pred. No. 65;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCACGCCAGGCATG 24
|||||
Db 35 AAAAAAAAAATCACGCCAGGCAG 12

RESULT 10
AW554415
LOCUS
DEFINITION
L0240C02-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
Clone L0240C02 3', mRNA sequence.
AW554415
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 254)
Tanaka,T.S., Jaradat,S.A., Lim,M.X., Kartul,G.J., Wang,X.,
Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,
Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
```

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 413)

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akınret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other\_GSSs: RPCI-23-375L24.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@igir.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page: [http://www.tigr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)

Plate: 375 row: L column: 24

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..413

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-375L24"

/sex="Female"

/lab\_host="DH10B"

/clone\_lib="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

Query Match 89.6%; Score 22.4; DB 9; Length 413;

Best Local Similarity 95.8%; Pred. No. 65;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCACGCCAGGCATG 24

|||||

Db 35 AAAAAAAAAATCACGCCAGGCAG 12

RESULT 10

AW554415

LOCUS

DEFINITION

L0240C02-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA

Clone L0240C02 3', mRNA sequence.

AW554415

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 254)

Tanaka,T.S., Jaradat,S.A., Lim,M.X., Kartul,G.J., Wang,X.,

Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,

Wood,W.H. III, Becker,K.G. and Ko,M.S.H.

Genome-wide expression profiling of mid-gestation placenta and

embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

Mon Feb 6 12:23:11 2006

PUBMED 10922068  
COMMENT On Mar 7, 2000 this sequence version replaced gi:7199826.

Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgaun.grc.nia.nih.gov  
Plate: L0240 row: C column: 02  
Seq primer: -21M13 Forward  
High quality sequence stop: 254  
POLYA=Yes.

FEATURES  
source Location/Qualifiers  
1. 254  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nia:ST:L0240C02-3"  
/db\_xref="taxon:10090"  
/clone="L0240C02"  
/sex="female"  
/dev\_stage="Newborn Ovary"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Newborn Ovary cDNA Library"  
/notes="vector: pSPORT1 (Gibco/BRL Life Technology);  
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from  
7 Newborn Ovary. The double-stranded cDNA was synthesized  
by Gibco's kit with an Oligo(dT) primer [NotI  
primer-adaptor from GibcoBRL]  
[5'-pGACTAGTCTAGATCGGAGCGCCCTTTT-3'] from  
2.56ug of total RNA. The double-stranded cDNAs were  
treated with T4 DNA polymerase and purified by  
ethanol-precipitation. The cDNAs were ligated to  
Lone-linker Li-Sal3 (include SalI sequence). The cDNAs  
were purified by phenol/chloroform and separated from  
free linkers by Centricon 100. Then, cDNAs were amplified  
by long-range high fidelity PCR using Takara's Ex Tag  
polymerase. Then, the cDNAs were purified by  
phenol/chloroform and by Centricon 100. The cDNAs were  
digested with SalI and NotI enzymes. Then, the cDNAs were  
size selected by Gibco's Size Fractionation Column. The  
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by chemical method. The library was  
constructed by Xiaohong Wang and Yulan Piao."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 1; Length 254;  
Best Local Similarity 92.0%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCAGCGCCGCGCATGG 25  
|||||  
Db 127 AAAAAAAAAATCTTGGCCGCGCATGG 151

RESULT 11  
BF469089/c  
LOCUS

DEFINITION BF469089 UI-M-BH3-atg-h-06-0-UI.r1 NIH BMAP\_M S4 Mus musculus cDNA clone  
UI-M-BH3-atg-h-06-0-UI 5', mRNA sequence.

ACCESSION BF469089

VERSION BF469089.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 317)

Bonaldi, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
should be noted that Bento Soares is generating a small number of  
additional specialized non-redundant arrays of BMAP cDNAs whose  
availability will be considered under appropriate and limited  
collaborative arrangements. The following repetitive elements were  
found in this cDNA sequence: 1-82, >PB1D10\$INE/Alu 87-108,  
>AT rich#Low complexity 344-364, >AT rich#Low complexity

Seq primer: M13 Reverse.

FEATURES  
source Location/Qualifiers

1. 317  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-atg-h-06-0-UI"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BMAP M S4"  
/notes="vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH BMAP\_M S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH BMAP M S4,  
NIH BMAP M S3.3, NIH BMAP M S3.2, NIH BMAP M S3.1,  
NIH BMAP\_M S2, NIH BMAP\_M S1. The subtracted library  
(NIH BMAP\_M S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH BMAP\_M S3.3, NIH BMAP\_M S3.2, and  
NIH BMAP\_M S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH BMAP M S3.3, NIH BMAP M S3.2, and NIH BMAP M S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH BMAP\_M S4 library. This procedure has been previously  
described [Bonaldi, Lennon and Soares, Genome Research  
6:791-806, 1996]"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 317;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCAGCGCCGCGCATGG 25

|||||  
Db 47 AAAAAAAAAATCTTGGCCGCGCATGG 23

RESULT 12

BB105780/c

LOCUS

DEFINITION

BB105780

322 bp mRNA linear EST 27-JUN-2000

between diaphragm region and neck Mus musculus cDNA clone

9430096N15 3' similar to L42293 Mus musculus acyl-coenzyme

A:cholesterol acyltransferase (ACACT) mRNA, mRNA sequence.

BB105780

ACCESSION

BB105780.1

VERSION

GI:8758348







purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centrion 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centrion 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pQWV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.7 kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 7; Length 330;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||||  
Db 256 AAAAAAAAAATCTGGCCAGGCATGG 280

## RESULT 14

BY690105/c  
LOCUS BY690105 372 bp mRNA linear EST 16-DEC-2002  
DEFINITION Musculus full-length enriched, osteoclast-like cell Mus  
musculus cDNA clone I420009013 3', mRNA sequence.

ACCESSION BY690105  
VERSION 1  
KEYWORDS GI:27097427

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalia, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L. G., Wainwright, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by Takashi Ishikawa (Department of Surgery

2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama

236-0004 Japan) whose assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

## FEATURES

## source

1..372  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I420009013"  
/cell\_type="osteoclast-like cell"  
/clone\_lib="RIKEN full-length enriched, osteoclast-like  
cell"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 5; Length 372;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||||  
Db 124 AAAAAAAAAATCTGGCCAGGCATGG 100

## RESULT 15

AW492561 373 bp mRNA linear EST 24-FEB-2000

LOCUS UT-M-BH3-atg-h-06-0-UI.s1 NIH BMAP M S4 Mus musculus cDNA clone

DEFINITION UT-M-BH3-atg-h-06-0-UI 3', mRNA sequence.

ACCESSION AW492561

VERSION AW492561.1 GI:7062842

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 373)

1 (bases 1 to 373)

Bonaldo, M. F., Lennon, G. and Soares, M. B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEST@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized brain stems library cDNA library Preparation: M.B. Soares lab clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 256-277, >AT-rich#Low complexity 282-363, >PB1D10#SINE/Alu  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

Location/Qualifiers  
 1. .373

source

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-atg-h-06-0-UI"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NIH BMAP M\_S4"  
 /note="Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BMAP M\_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH BMAP M\_S4, NIH BMAP M\_S3.3, NIH BMAP M\_S3.2, NIH BMAP M\_S1, NIH BMAP M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH BMAP M\_S3.3, NIH BMAP M\_S3.2, and NIH BMAP M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH BMAP M\_S3.3, NIH BMAP M\_S3.2, and NIH BMAP M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH BMAP M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-805, 1996)  
 TAG\_TISSUE=brain-stems  
 TAG\_LIB=NIH\_BMAP\_M\_S4  
 TAG\_SEQ=TCATG"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 1; Length 373;  
 Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
 Db 268 AAAAAAAAAATCTTGGCCAGCATGG 292

## RESULT 16

BB791462 375 bp mRNA linear EST 08-JUL-2003  
 LOCUS BB791462 RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA Mus  
 DEFINITION

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

musculus cDNA clone G430136J01 3', mRNA sequence.

BB791462  
 BB791462.1 GI:16959958

EST.

Mus musculus (house mouse)

ORGANISM

Mus musculus

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

1 (bases 1 to 375)  
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Makuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

## TITLE

Unpublished (2001)

## JOURNAL

## COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

## FEATURES

source

Location/Qualifiers

1. .375  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="G430136J01"  
 /cell\_line="CRL-1751 WEHI 164"  
 /clone\_lib="RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 375;  
 Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCAGCGCCAGCATGG 25  
 Db 121 AAAAAAAAAATCTTGGCCAGCATGG 97  
 RESULT 17  
 BY698909/c

**LOCUS** BY698909 379 bp mRNA linear EST 16-DBC-2002  
**DEFINITION** BY698909 RIKEN full-length enriched, osteoclast-like cell Mus musculus cDNA clone 1420113M05 3', mRNA sequence.  
**ACCESSION** BY698909  
**VERSION** BY698909.1 GI:27109934  
**KEYWORDS** EST.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
**AUTHORS** Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojohori, T., Balderelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.E., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Saitelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sundland, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomika, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.  
**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**PUBLISHED** 12466851  
**COMMENT** Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.: Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Takashi Ishikawa ( Department of Surgery 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan ) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES** Location/Qualifiers  
 1..379  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strains="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="1420113M05"  
 /cell\_type="osteoclast-like cell"  
 /clone\_lib="RIKEN full-length enriched, osteoclast-like cell"  
**ORIGIN**  
 Query Match 87.2%; Score 21.8; DB 5; Length 379;  
 Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAATCACGGCCAGGCATGG 25  
 |||||  
 Db 126 AAAAAAAAAATCTGGCCAGGCATGG 102  
 |||||  
**RESULT** 18  
 CN698559 366 bp mRNA linear EST 18-MAY-2004  
**LOCUS** E0409E05-5 NIA Mouse E11.5 whole embryo cDNA library (Long) Mus musculus cDNA clone NIA:E0409E05 IMAGE:30869428 5', mRNA sequence.  
**DEFINITION** CN698559  
**ACCESSION** CN698559.1 GI:47467308  
**VERSION** EST.  
**KEYWORDS** Mus musculus (house mouse)  
**SOURCE** Mus musculus  
**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 386)  
**AUTHORS** Sharov, A.A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y., VanBuren, V., Falco, G., Martin, P.R., Stagg, C.A., Basse, U.C., Wang, Y., Carter, M.G., Hamatani, T., Alba, K., Akutsu, H., Sharova, L., Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S., Nagaraja, R., Boheler, K.R., Taub, D., Hodes, R.J., Longo, D.L., Schlessinger, D., Kellier, J., Klotz, E., Kelsoe, G., Umezawa, A., Vescovi, A.L., Rossant, J., Kunath, T., Hogan, B.L., Curci, A., D'Urso, M., Kelso, J., Hide, W. and Ko, M.S.  
 Transcriptome analysis of mouse stem cells and early embryos  
 PLOS Biol. 1 (3), 410-419 (2003)  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Casseell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: [cdna@leuon.gsc.nia.nih.gov](mailto:cdna@leuon.gsc.nia.nih.gov)  
 Plate: E0409 row: E column: 05  
 Seq primer: M13 Reverse  
 High quality sequence stop: 386  
 POLYA=No.  
**FEATURES** Location/Qualifiers  
 1..386  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strains="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0409E05 IMAGE:30869428"  
 /tissue\_type="whole embryo including extraembryonic tissues at 11.5-days postcoitum"  
 /dev\_stage="E11.5"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse E11.5 whole embryo cDNA library (Long)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 3 embryos at 11.5-days postcoitum. Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen): 5'-pGACTAGTCTAGATCGCGAGCGCCCTCTTTTTTTTTT-3' from 2ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.3Kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 87.2%; Score 21.8; DB 7; Length 386;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 255 AAAAAAAAAATCTGGCCAGGCATGG 279

RESULT 19

BY425365/C

LOCUS

DEFINITION

musculus cDNA clone 1920078L02 3', mRNA sequence.

BY425365

EST.

Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

BY425365 388 bp mRNA linear EST 13-DEC-2002  
BY425365 RIKEN full-length enriched, 16 days embryo kidney Mus  
musculus cDNA clone 1920078L02 3', mRNA sequence.  
BY425365  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
1 (bases 1 to 388)  
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaido,I., Oatato,N., Saito,R., Suzuki,H., Yamanaka,I.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,  
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,  
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusci,V.,  
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,  
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,  
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,  
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,  
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,  
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G.,  
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,  
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,  
Santelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,  
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,  
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,  
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,  
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Inotani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,

Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL:<http://genome.gsc.riken.jp/>  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

FEATURES  
source

1..388  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1920078L02"  
/tissue\_type="kidney"  
/dev\_stage="16 days embryo"  
/clone\_lib="RIKEN full-length enriched, 16 days embryo  
kidney"

ORIGIN

Query Match 87.2%; Score 21.8; DB 5; Length 388;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 136 AAAAAAAAAATCTGGCCAGGCATGG 112

RESULT 20

BY684912/C

LOCUS

DEFINITION

clone 1320008M16 3', mRNA sequence.

BY684912

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

BY684912 392 bp mRNA linear EST 16-DEC-2002  
BY684912 RIKEN full-length enriched, stroma cell Mus musculus cDNA  
clone 1320008M16 3', mRNA sequence.  
BY684912  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE AUTHORS	1 (bases 1 to 392) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusici, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Jarvis, E. D., Kanai, A., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL PUBLISHED	of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
COMMENT	12466851 Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@genome.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Takashi Ishikawa (Department of Surgery 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan) whose assistance we gratefully acknowledge. Please visit our web site (http://genome.gsc.riken.go.jp) for further details Location/Qualifiers 1. .392 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090"
FEATURES Source	Location/Qualifiers 1. .392 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090"
ORIGIN	Query Match 87.2%; Score 21.8; DB 5; Length 392; Best Local Similarity 92.0%; Pred. No. 1.2e+02; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0; /clone="I320008M16" /cell_type="stroma cell" /clone_lib="RIKEN full-length enriched, stroma cell"
QY	1 AAAAAAAAAATCAGCGCCAGCGCATGG 25 
Db	134 AAAAAAAAAATCTGGCCAGCGCATGG 110 
RESULT 21	
LOCUS	BY691014 394 bp mRNA linear EST 16-DEC-2002
DEFINITION	BY691014 RIKEN full-length enriched, osteoclast-like cell Mus musculus cDNA clone 1420015L11 3', mRNA sequence.
ACCESSION	BY691014
VERSION	BY691014.1 GI:27099180
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Murinae; Mus. 1 (bases 1 to 394)
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusici, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Jarvis, E. D., Kanai, A., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Tissues were provided by Takashi Ishikawa ( Department of Surgery 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan ) whose assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

```

1. .394
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I420015L11"
/cell_type="osteoclast-like cell"
/clone_lib="RIKEN full-length enriched, osteoclast-like cell"

```

## ORIGIN

```

Query Match      87.2%; Score 21.8; DB 5; Length 394;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAAAAAAAAATCAGCGCAGGATGG 25

Db 154 AAAAAAAAAATCTGGCCAGGATGG 130

## RESULT 22

BY701163/c

LOCUS

```

DEFINITION
Musculus cDNA clone I420130B21 3', mRNA sequence.

```

ACCESSION

BY701163

VERSION

EST.

Mus musculus

Mus musculus

ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

```

REFERENCE

AUTHORS

```

1 (bases 1 to 401)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.N., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Sempie,C.A., Setou,M., Shimada,K.,

```

Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,  
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,  
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,  
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,K., Shibata,K.,  
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
 Rogers,J., Birney,E. and Hayashizaki,Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 12466851  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL:<http://genome.gsc.riken.jp/>  
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
 Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
 Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
 Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
 Shiraki,T., Tagami,M., Waki,K., Watanabe,A., Muranatsu,M. and  
 Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Takashi Ishikawa ( Department of Surgery  
 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama  
 236-0004 Japan ) whose assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

## FEATURES

source

```

1. .401
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I420130B21"
/cell_type="osteoclast-like cell"
/clone_lib="RIKEN full-length enriched, osteoclast-like cell"

```

## ORIGIN

```

Query Match      87.2%; Score 21.8; DB 5; Length 401;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAAAAAAAAATCAGCGCAGGATGG 25

Db 141 AAAAAAAAAATCTGGCCAGGATGG 117

## RESULT 23

BY694371/c

LOCUS

```

DEFINITION
Musculus cDNA clone I420130B21 3', mRNA sequence.

```

```

ACCESSION
BY694371
VERSION
EST.
Mus musculus
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 401)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.N., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Sempie,C.A., Setou,M., Shimada,K.,

```



DEFINITION	BY694371 RIKEN full-length enriched, osteoclast-like cell Mus musculus cDNA clone 1420033A09 3', mRNA sequence.
ACCESSION	BY694371
VERSION	BY694371.1 GI:27105326
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Chordata; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 403) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.D., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
PUBLISHED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
FEATURES	source 1. .403 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="taxon:10090" /clone="1420033A09" /cell_type="osteoclast-like cell" /clone_lib="RIKEN full-length enriched, osteoclast-like cell"
ORIGIN	Query Match 87.2%; Score 21.8; DB 5; Length 403; Best Local Similarity 92.0%; Pred. No. 1.2e+02; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0; QY 1 AAAAAAAAAATCAGCGCAGGCATGG 25 Db 146 AAAAAAAAAATCTGGCCAGGCATGG 122
RESULT 24	BY690774/c
LOCUS	BY690774
DEFINITION	Mus musculus
ACCESSION	BY690774
VERSION	BY690774.1 GI:27098712
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus. 1 (bases 1 to 404) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.D., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
PUBLISHED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

DEFINITION	BY694371 RIKEN full-length enriched, osteoclast-like cell Mus musculus cDNA clone 1420033A09 3', mRNA sequence.
ACCESSION	BY694371
VERSION	BY694371.1 GI:27105326
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Chordata; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 403) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.D., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
PUBLISHED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Takashi Ishikawa (Department of Surgery

2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan ) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

Location/Qualifiers  
1. .403  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="1420033A09"  
/cell\_type="osteoclast-like cell"  
/clone\_lib="RIKEN full-length enriched, osteoclast-like cell"

Query Match 87.2%; Score 21.8; DB 5; Length 403;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCAGGCATGG 25  
Db 146 AAAAAAAAAATCTGGCCAGGCATGG 122

RESULT 24  
BY690774/c

LOCUS  
BY690774

DEFINITION  
Mus musculus

ACCESSION  
BY690774

VERSION  
BY690774.1 GI:27098712

KEYWORDS  
EST.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 404)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.D., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL  
Nature 420, 563-573 (2002)

PUBLISHED  
12466851

COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute



The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Takaishi Ishikawa ( Department of Surgery  
2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama  
236-0004 Japan ) whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

Location/Qualifiers  
1. .404  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I420014G11"  
/cell\_type="osteoclast-like cell"  
/clone\_lib="RIKEN full-length enriched, osteoclast-like  
cell"

ORIGIN  
Query Match 87.2%; Score 21.8; DB 5; Length 404;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGGCGCAGGCATGG 25  
145 AAAAAAAAAATCTGGCCAGGCATGG 121

RESULT 25  
BY655013/c  
LOCUS  
DEFINITION  
cDNA clone K530321K23 3', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 415)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,

Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pettea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
Ravasi, F., Reed, J.C., Reid, D.J., Reid, J., King, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Vetardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michela Fagiolini and Takao K. Hensch (   
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan ) whose  
assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

Location/Qualifiers  
1. .415  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K530321K23"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"  
FEATURES  
source  
Query Match 87.2%; Score 21.8; DB 5; Length 415;  
ORIGIN

Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 156 AAAAAAAAAATCTGGCCAGGCATGG 132

RESULT 26  
BG074809  
LOCUS  
DEFINITION  
H3139F01-3 N1A Mouse 15K cDNA Clone Set Mus musculus cDNA clone

ACCESSION  
BG074809  
VERSION  
KEYWORDS  
SOURCE  
EST.

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X.,  
Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H.,  
Wobd, W.H., III, Becker, K.G. and Ko, M.S.H.

TITLE  
Genome-wide expression profiling of mid-gestation placenta and  
embryo using a 15,000 mouse developmental cDNA microarray  
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

JOURNAL  
PUBMED  
COMMENT  
On Jan 26, 2001 this sequence version replaced gi:12557378.  
Other ESTs: H3139F01-5  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit <http://lgsun.grc.nia.nih.gov/cdna/15k.html> for details.  
Plate: H3139 row: F column: 01  
Seq primer: -21M13 Forward  
High quality sequence stop: 418  
POLYA=Yes.

FEATURES  
source  
Location/Qualifiers

1. 418  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:H3139F01-3"  
/db\_xref="taxon:10090"  
/clone="H3139F01"  
/sex="Clones arrayed from a variety of cDNA libraries"  
/dev\_stage="Clones arrayed from a variety of cDNA libraries"  
/lab\_host="DH10B"  
/clone\_lib="N1A Mouse 15K cDNA Clone Set"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This  
clone is among a rearranged set of 15,247 clones from 11  
embryo cDNA libraries (including preimplantation stage  
embryos from unfertilized egg to blastocyst, embryonic  
part of E7.5 embryos, extraembryonic part of E7.5  
embryos, and E12.5 female mesonephros/gonad) and one  
newborn ovary cDNA library. Average insert size 1.5 kb.  
All source libraries are cloned unidirectionally with  
Oligo(dT)-Not primers. References include: (1)  
Genome-wide expression profiling of mid-gestation  
placenta and embryo using a 15,000 mouse developmental  
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97:  
9127-9132; (2) Large-scale cDNA analysis reveals phased  
gene expression patterns during preimplantation mouse  
development, 2000, Development, 127: 1737-1749; (3)  
Genome-wide mapping of unselected transcripts from  
extraembryonic tissue of 7.5-day mouse embryos reveals  
enrichment in the t-complex and under-representation on

the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

Query Match 87.2%; Score 21.8; DB 2; Length 418;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 254 AAAAAAAAAATCTGGCCAGGCATGG 278

RESULT 27  
W12097/c  
LOCUS

DEFINITION  
W12097  
ma96f07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
IMAGE:318565 5', mRNA sequence.

ACCESSION  
W12097  
VERSION  
KEYWORDS  
SOURCE  
EST.

Mus musculus (house mouse)  
Mus musculus

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE  
The WashU-HMI Mouse EST Project  
JOURNAL  
COMMENT  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI: 209181

Seq primer: mob.REGA+ET  
High quality sequence stop: 176.  
Location/Qualifiers

FEATURES  
source  
Location/Qualifiers

1. 418  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:318565"  
/dev\_stage="19.5 dpc total fetus"  
/clone\_lib="Soares mouse p3NMF19.5"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 418;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||

Db 161 AAAAAAAAAATCTTGGCCAGGCATGG 137

RESULT 28  
 A1124789/c  
 LOCUS  
 DEFINITION  
 A1124789 421 bp mRNA linear EST 11-SEP-1998  
 an56910.x1 Johnston frontal cortex Homo sapiens cDNA clone  
 IMAGE:1539547 3' similar to contains Alu repetitive element;; mRNA  
 sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)

REFERENCE  
 AUTHORS  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -49m13 fwd. BT from Amersham.

FEATURES  
 source  
 1. 421  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1539547"  
 /sex="male"  
 /tissue\_type="pooled frontal lobe"  
 /dev\_stage="adult"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Johnston frontal cortex"  
 /note="Organ: brain; Vector: Bluescript SK-; Site\_1:  
 EcoRI; Stanley Neurophatology Consortium  
 (www.stanleylab.org) brains S-58, S-65, S-67, S-78.  
 Random + oligo-dT primed into EcoRI site of ZAP II Vector.  
 Mass excised. Avg insert length 1.9kb. Custom library  
 provided by Dr. Nancy Johnston [(410) 614-3918,  
 nlj@welchlink.welch.jhu.edu]."

ORIGIN  
 Query Match 87.2%; Score 21.8; DB 1; Length 421;  
 Best Local Similarity 92.0%; Pred. NO. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCAGGCATGG 25  
 |||||  
 |||||

Db 283 AAAAAAAAAATTACGCCAGGCATAG 259  
 |||||  
 |||||

RESULT 29  
 BY697032/c  
 LOCUS  
 DEFINITION  
 BY697032 423 bp mRNA linear EST 16-DEC-2002  
 musculus cDNA clone I420102C01 3', mRNA sequence.  
 BY697032  
 BY697032.1 GI:27108028  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 421)  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -49m13 fwd. BT from Amersham.

FEATURES  
 source  
 1. 421  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1539547"  
 /sex="male"  
 /tissue\_type="pooled frontal lobe"  
 /dev\_stage="adult"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Johnston frontal cortex"  
 /note="Organ: brain; Vector: Bluescript SK-; Site\_1:  
 EcoRI; Stanley Neurophatology Consortium  
 (www.stanleylab.org) brains S-58, S-65, S-67, S-78.  
 Random + oligo-dT primed into EcoRI site of ZAP II Vector.  
 Mass excised. Avg insert length 1.9kb. Custom library  
 provided by Dr. Nancy Johnston [(410) 614-3918,  
 nlj@welchlink.welch.jhu.edu]."

ORIGIN  
 Query Match 87.2%; Score 21.8; DB 1; Length 421;  
 Best Local Similarity 92.0%; Pred. NO. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCAGGCATGG 25  
 |||||  
 |||||

Db 283 AAAAAAAAAATTACGCCAGGCATAG 259  
 |||||  
 |||||

RESULT 29  
 BY697032/c  
 LOCUS  
 DEFINITION  
 BY697032 423 bp mRNA linear EST 16-DEC-2002  
 musculus cDNA clone I420102C01 3', mRNA sequence.  
 BY697032  
 BY697032.1 GI:27108028  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 421)  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -49m13 fwd. BT from Amersham.

FEATURES  
 source  
 1. 421  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1539547"  
 /sex="male"  
 /tissue\_type="pooled frontal lobe"  
 /dev\_stage="adult"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Johnston frontal cortex"  
 /note="Organ: brain; Vector: Bluescript SK-; Site\_1:  
 EcoRI; Stanley Neurophatology Consortium  
 (www.stanleylab.org) brains S-58, S-65, S-67, S-78.  
 Random + oligo-dT primed into EcoRI site of ZAP II Vector.  
 Mass excised. Avg insert length 1.9kb. Custom library  
 provided by Dr. Nancy Johnston [(410) 614-3918,  
 nlj@welchlink.welch.jhu.edu]."

ORIGIN  
 Query Match 87.2%; Score 21.8; DB 1; Length 421;  
 Best Local Similarity 92.0%; Pred. NO. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCAGGCATGG 25  
 |||||  
 |||||

Db 283 AAAAAAAAAATTACGCCAGGCATAG 259  
 |||||  
 |||||

RESULT 29  
 BY697032/c  
 LOCUS  
 DEFINITION  
 BY697032 423 bp mRNA linear EST 16-DEC-2002  
 musculus cDNA clone I420102C01 3', mRNA sequence.  
 BY697032  
 BY697032.1 GI:27108028  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 421)  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -49m13 fwd. BT from Amersham.

FEATURES  
 source  
 1. 421  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1539547"  
 /sex="male"  
 /tissue\_type="pooled frontal lobe"  
 /dev\_stage="adult"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Johnston frontal cortex"  
 /note="Organ: brain; Vector: Bluescript SK-; Site\_1:  
 EcoRI; Stanley Neurophatology Consortium  
 (www.stanleylab.org) brains S-58, S-65, S-67, S-78.  
 Random + oligo-dT primed into EcoRI site of ZAP II Vector.  
 Mass excised. Avg insert length 1.9kb. Custom library  
 provided by Dr. Nancy Johnston [(410) 614-3918,  
 nlj@welchlink.welch.jhu.edu]."

ORIGIN  
 Query Match 87.2%; Score 21.8; DB 1; Length 421;  
 Best Local Similarity 92.0%; Pred. NO. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCAGGCATGG 25  
 |||||  
 |||||

Db 283 AAAAAAAAAATTACGCCAGGCATAG 259  
 |||||  
 |||||

RESULT 29  
 BY697032/c  
 LOCUS  
 DEFINITION  
 BY697032 423 bp

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I420102C01"
/cell_type="osteoclast-like cell"
/clone_lib="RIKEN full-length enriched, osteoclast-like
cell"

ORIGIN

Query Match      87.2%; Score 21.8; DB 5; Length 423;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    |||||
Db 176 AAAAAAAAAATCTGGCCAGGCATGG 152

RESULT 30
BB781154/c
LOCUS
DEFINITION BB781154 425 bp mRNA linear EST 08-JUL-2003
13.20-3B3 cDNA Mus musculus cDNA clone G430071G05 3', mRNA
sequence.
BB781154
BB781154.1 GI:16941854
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM

REFERENCE
AUTHORS
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hirozane,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numaaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Tanaka,A., Tanaka,T., Tanaka,T.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

TITLE
Unpublished (2001)
JOURNAL
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues. Location/Qualifiers

```

```

source
1. 425
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="G430071G05"
/cell_type="B cells"
/clone_lib="RIKEN full-length enriched, B cells CRL-1669
BCL1 clone 13.20-3B3 cDNA"

ORIGIN

Query Match      87.2%; Score 21.8; DB 2; Length 425;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    |||||
Db 167 AAAAAAAAAATCTTGGCCAGGCATGG 143

RESULT 31
BB789804/c
LOCUS
DEFINITION BB789804 425 bp mRNA linear EST 08-JUL-2003
Mus musculus cDNA clone G430129E01 3', mRNA sequence.
BB789804
BB789804.1 GI:16958300
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM

REFERENCE
AUTHORS
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hirozane,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numaaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Tanaka,A., Tanaka,T., Tanaka,T.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

TITLE
Unpublished (2001)
JOURNAL
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues. Location/Qualifiers

```

further details.

FEATURES  
source

Location/Qualifiers  
1. .425  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="G430129E01"  
/cell\_line="RCB-0549 Cle-H3"  
/dev\_stage="embryo"  
/clone\_lib="RIKEN full-length enriched, embryo RCB-0549 Cle-H3-CDNA"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 425;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 170 AAAAAAAAAATCTGGCCAGGCATGG 146

## RESULT 32

BY383430/c  
LOCUS BY383430 RIKEN full-length enriched, 4 days embryo whole body Mus  
musculus cDNA clone IC0039E02 3', mRNA sequence.  
EST. BY383430.1 GI:26612918

## ACCESSION

BY383430

## KEYWORDS

Mus musculus (house mouse)

## SOURCE

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 426)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batilov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reed, J.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandalin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L.G., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct SubmissionComputational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.Location/Qualifiers  
1. .426/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IC0039E02"  
/tissue\_type="whole body"  
/dev\_stage="4 days embryo"  
/clone\_lib="RIKEN full-length enriched, 4 days embryo  
whole body"

Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 664 Std Error: 0.00  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 424.  
Location/Qualifiers

# FEATURES

source

```
1. .431
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1283781"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP GCB1"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

# ORIGIN

```
Query Match      87.2%; Score 21.8; DB 1; Length 431;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAACAGCGCCAGCATGG 25
    |||||
Db 431 AAAAAAAAAATCCAGCGCGCATGG 407
```

```
RESULT 34
A1642924
LOCUS
DEFINITION
A1642924 431 bp mRNA linear EST 29-APR-1999
v652a08.x1 Knowles Solter mouse blastocyst B1 Mus musculus
clone IMAGE:805142 3' similar to gb:L42293 Mus musculus
acyl-coenzyme A:cholesterol acyltransferase (MOUSE); mRNA
sequence.
```

```
ACCESSION
A1642924
VERSION
A1642924.1 GI:4721399
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
```

```
REFERENCE
AUTHORS
Marra,M., Hallier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Pardoll,M., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
```

```
TITLE
JOURNAL
COMMENT
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
```

MGI:481486

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Seq primer: Primer name ambiguous  
High quality sequence stop: 422.  
Location/Qualifiers

# FEATURES

source

```
1. .431
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/db_xref="taxon:10090"
/clone="IMAGE:805142"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/clone_lib="Knowles Solter mouse blastocyst B1"
/notes="Organ: embryo; Vector: pSPORT; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI (dT):
5'-CGTCCAGCGTCGACGCTTTTTTTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
```

# ORIGIN

```
Query Match      87.2%; Score 21.8; DB 1; Length 431;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCAGCGCCAGCATGG 25
    |||||
Db 259 AAAAAAAAAATCTGGCCAGCATGG 283
```

# RESULT 35

BY693822/c

LOCUS

DEFINITION

```
BY693822 RIKEN full-length enriched, osteoclast-like cell Mus
musculus cDNA clone I42029M24 3', mRNA sequence.
```

```
ACCESSION
BY693822
VERSION
BY693822.1 GI:27104525
```

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
```

REFERENCE

AUTHORS

```
1 (bases 1 to 434)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Cojocari,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Redziarski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Santelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Alzawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
```



**TITLE**  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
**JOURNAL**  
Analysis of the mouse transcriptome based on functional annotation  
**PUBMED**  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
**COMMENT**  
12466851

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9212  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Takaashi Ishikawa ( Department of Surgery  
2 Yokohama City University 3-9 Fukura, Kanazawa-ku, Yokohama  
236-0004 Japan ) whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

# FEATURES

source  
1. .434  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I420029M24"  
/cell\_type="osteoclast-like cell"  
/clone\_lib="RIKEN full-length enriched, osteoclast-like  
cell"  
ORIGIN  
Query Match : 87.2%; Score 21.8; DB 5; Length 434;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 177 AAAAAAAAAATCTGGCCAGGCATGG 153  
|||||

RESULT 36  
BY384796/c  
LOCUS  
DEFINITION  
BY384796 436 bp mRNA linear EST 12-DEC-2002  
clone IIC0004N17 3', mRNA sequence.  
ACCESSION  
BY384796.1 GI:26614284  
VERSION  
BY384796  
KEYWORDS  
Mus musculus (house mouse)  
SOURCE  
Mus musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
1 (bases 1 to 436)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojibori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chotha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Flatcher, C., Frazer, K.S., Gaasterland, T.,  
Gustinich, S., Gliori, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, T., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851

**TITLE**  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
**JOURNAL**  
cDNA library was prepared and sequenced in Mouse Genome  
**PUBMED**  
Encyclopedia Project of Genome Exploration Research Group in Riken  
**COMMENT**  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Takaashi Ishikawa ( Department of Surgery  
2 Yokohama City University 3-9 Fukura, Kanazawa-ku, Yokohama  
236-0004 Japan ) whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

Location/Qualifiers  
1. .436  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IIC0004N17"  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

# FEATURES

Location/Qualifiers  
1. .436  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IIC0004N17"



/dev stage="blastocyst"  
/clone\_lib="RIKEN full-length enriched, blastocyst"

ORIGIN  
Query Match 87.2%; Score 21.8; DB 5; Length 436;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCGGCCAGCATGG 25  
|||||||  
DB 180 AAAAAAAAAATCGGCCAGCATGG 156

RESULT 37  
BB782484/c  
LOCUS BB782484 RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA Mus  
DEFINITION BB782484  
ACCESSION BB782484.1 GI:16943184  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 447)  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
MURAMATSU, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

FEATURES Location/Qualifiers  
1. .447  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/db\_xref="taxon:10090"

/clone="G430081B13"  
/cell\_line="CRL-1751 WEHI 164"  
/clone\_lib="RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA"

ORIGIN  
Query Match 87.2%; Score 21.8; DB 2; Length 447;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCGGCCAGCATGG 25  
|||||||  
DB 193 AAAAAAAAAATCGGCCAGCATGG 169

RESULT 38  
BB819202/c  
LOCUS BB819202 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus  
DEFINITION BB819202  
ACCESSION BB819202.1 GI:16991831  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 470)  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
MURAMATSU, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

FEATURES Location/Qualifiers  
1. .470  
/organism="Mus musculus"  
/mol\_type="mRNA"

purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lU-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 474;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 253 AAAAAAAAAATCTTGGCCAGGCATGG 277

## RESULT 40

AZ450659/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

/db\_xref="taxon:10090"  
/clone="G730048H03"  
/tissue\_type="lung"  
/cell\_line="RCB-0558 LLC"  
/clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC  
cDNA"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 470;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 218 AAAAAAAAAATCTTGGCCAGGCATGG 194

## RESULT 39

BM205156

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

COMMENT

## FEATURES

source

BM205156 474 bp mRNA linear EST 08-JUN-2003  
C0279F09-3 NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long) Mus  
musculus cDNA clone NIA:C0279F09 IMAGE:30018212 3', mRNA sequence.

BM205156

EST.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 474)

Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.

Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

Genome Res. 11 (9), 1553-1558 (2001)

11541199

On Dec 14, 2001 this sequence version replaced gi:17760224.

Other ESTs: C0279F09-5N

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@nsls.nih.gov

Plate: C0279 row: F column: 09

Seq primer: -21M13 Forward

High quality sequence stop: 474

POLYA=Yes

Location/Qualifiers

1..474

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="niaEST:C0279F09-3"

/db\_xref="taxon:10090"

/clone="NIA:C0279F09 IMAGE:30018212"

/tissue\_type="whole embryo including extraembryonic

tissues at 7.5-days postcoitum"

/dev\_stage="7.5-days postcoitum"

/lab\_host="DH10B"

/clone\_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library

(Long)"

/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (http://nsls.nih.gov/cDNA). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were

extracted from a pool of four embryos at 7.5-days

postcoitum. Double-stranded cDNAs were synthesized with an

oligo(dT) primer [Invitrogen:

5'-pGACTAGTCTAGATCGAGCGCCCTTTT-3'] from

7 ug of total RNA, treated with T4 DNA polymerase, and

purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lU-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 474;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

Db 253 AAAAAAAAAATCTTGGCCAGGCATGG 277

## RESULT 40

AZ450659/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

source

AZ450659 491 bp DNA linear GSS 04-OCT-2000  
IM0249D01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0249D01 R, genomic survey sequence.

AZ450659

GSS.

AZ450659.1 GI:10605669

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 491)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0249 row: D column: 01

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 491.

Location/Qualifiers

1..491

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0249D01"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PMW42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 491;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 120 AAAAAAAAAATCTGGCCAGGCATGG 96  
|||||

## RESULT 41

BM203574 514 bp mRNA linear EST 08-JUN-2003  
LOCUS C0256E04-3 NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long) Mus  
DEFINITION musculus cDNA clone NIA:C0256E04 IMAGE:30015987 3', mRNA sequence.

ACCESSION BM203574 GI:31528419

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 514)

Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.

Construction of long-transcript enriched cDNA libraries from  
submicrogram amounts of total RNAs by a universal PCR amplification  
method

Genome Res. 11 (9), 1553-1558 (2001)

1154199

On Dec 14, 2001 this sequence version replaced gi:17758197.

Other ESTs: C0256E04-5N

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

Plate: C0256 row: E column: 04

Seq primer: -21M13 Forward

High quality sequence stop: 514

POIYA=Yes

Location/Qualifiers

## FEATURES

source

1..514  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:C0256E04-3"  
/db\_xref="taxon:10090"  
/clone="NIA:C0256E04 IMAGE:30015987"  
/tissue\_type="whole embryo including extraembryonic  
tissues at 7.5-days postcoitum"  
/dev\_stage="7.5-days postcoitum"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library  
(Long)"  
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:  
NotI; Mouse cDNA project by the Laboratory of Genetics,  
National Institute on Aging (NIA), Intramural Research  
Program, NIH (<http://lgsun.grc.nia.nih.gov/cdna>). This is

a long-transcript enriched cDNA library (Ref. Genome Res.  
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were  
extracted from a pool of four embryos at 7.5-days  
postcoitum. Double-stranded cDNAs were synthesized with an  
Oligo(dT) primer [Invitrogen].

5'-pactagttctagatcgacggcgccctttttttttt-3') from  
7 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to Lona-linker LL-Sal4, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
The DH10B *E. coli* host was transformed with the ligation  
mixture by the standard chemical method. The average  
insert size is about 2.2 kb. The library was constructed  
by Yulan Piao (NIA)."

## ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 514;  
Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 252 AAAAAAAAAATCTGGCCAGGCATGG 276  
|||||

## RESULT 42

AA721682/c

LOCUS AA721682

DEFINITION

AA721682 519 bp mRNA linear EST 18-FEB-1998  
ny72h09 s1 NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1283873 3'  
similar to TR:000549 000549 ORF2-LIKE PROTEIN ; contains Alu  
repetitive element; contains L1.t2 L1 repetitive element ;, mRNA  
sequence.

ACCESSION AA721682

VERSION AA721682.1 GI:2736592

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

1 (bases 1 to 519)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/IMG at:

[www-bio.llnl.gov/bbr/image/image.html](http://www-bio.llnl.gov/bbr/image/image.html)

Insert Length: 894 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 440.

## FEATURES

source

1..519  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1283873"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_GCB1"



Db 312 AAAAAAAAAATCTTGGCAGGCATGG 336

RESULT 45  
 CK345513 593 bp mRNA linear EST 22-DEC-2003  
 LOCUS L0247D01-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA  
 DEFINITION Clone L0247D01 3', mRNA sequence.  
 CK345513  
 ACCESSION CK345513  
 VERSION CK345513.1 GI:40301126  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 593)  
 AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.  
 TITLE Construction of long-transcript enriched cDNA libraries from  
 submicrogram amounts of total RNAs by a universal PCR amplification  
 method  
 JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
 PUBMED 11544199  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: L0247 row: D column: 01  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 593  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1..593  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="niaEST:L0247D01-3"  
 /db\_xref="taxon:10090"  
 /clone="L0247D01"  
 /sex="female"  
 /dev\_stage="Newborn Ovary"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Newborn Ovary cDNA Library"  
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology);  
 Site 1: SalI; Site 2: NotI; Total RNAs were extracted from  
 7 Newborn Ovary. The double-stranded cDNA was synthesized  
 by Gibco's kit with an Oligo(dT) primer [NotI  
 primer-adaptor from GibcoBRL]  
 [5'-pGACTAGTTCTAGATCGGCGCGCCCTTTT-3'] from  
 2.56ug of total RNA. The double-stranded cDNAs were  
 treated with T4 DNA polymerase and purified by  
 ethanol-precipitation. The cDNAs were ligated to  
 Lone-linker L1-Sal3 (include SalI sequence). The cDNAs  
 were purified by phenol/chloroform and separated from  
 free linkers by Centricon 100. Then, cDNAs were amplified  
 by long-range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centricon 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was  
 constructed by Xiaohong Wang and Yulan Piao."

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 7; Length 593;  
 Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAATCTTGGCAGGCATGG 25  
 ||||||||| |||||||||

Db 254 AAAAAAAAAATCTTGGCAGGCATGG 278

RESULT 46  
 AW554397 595 bp mRNA linear EST 31-AUG-2000  
 LOCUS L0240B05-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA  
 DEFINITION Clone L0240B05 3', mRNA sequence.  
 AW554397  
 ACCESSION AW554397.1 GI:7199820  
 VERSION AW554397.1  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 595)  
 AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,  
 Grabovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,  
 Wood,W.H. III, Becker,K.G. and Ko,M.S.H.  
 TITLE Genome-wide expression profiling of mid-gestation placenta and  
 embryo using a 15,000 mouse developmental cDNA microarray  
 Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
 JOURNAL 10922068  
 PUBMED  
 COMMENT Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: L0240 row: B column: 05  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 595  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1..595  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="niaEST:L0240B05-3"  
 /db\_xref="taxon:10090"  
 /clone="L0240B05"  
 /sex="female"  
 /dev\_stage="Newborn Ovary"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Newborn Ovary cDNA Library"  
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology);  
 Site 1: SalI; Site 2: NotI; Total RNAs were extracted from  
 7 Newborn Ovary. The double-stranded cDNA was synthesized  
 by Gibco's kit with an Oligo(dT) primer [NotI  
 primer-adaptor from GibcoBRL]  
 [5'-pGACTAGTTCTAGATCGGCGCGCCCTTTT-3'] from  
 2.56ug of total RNA. The double-stranded cDNAs were  
 treated with T4 DNA polymerase and purified by  
 ethanol-precipitation. The cDNAs were ligated to  
 Lone-linker L1-Sal3 (include SalI sequence). The cDNAs  
 were purified by phenol/chloroform and separated from  
 free linkers by Centricon 100. Then, cDNAs were amplified  
 by long-range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centricon 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was  
 constructed by Xiaohong Wang and Yulan Piao."

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 1; Length 595;  
 Best Local Similarity 92.0%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAATCTTGGCAGGCATGG 25



Search completed: February 3, 2006, 22:01:35  
Job time : 2963.67 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-1

Perfect score: 25

Sequence: 1 aaaaaaaatcacggcaggtgg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCPUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.4	93.6	51905	3	US-09-949-002-667
C 2	23.4	93.6	51905	3	US-09-949-002-781
C 3	23.4	93.6	107827	3	US-09-949-016-15790
C 4	21.8	87.2	23640	3	US-09-949-016-13646
C 5	21.8	87.2	27592	3	US-09-949-016-14682
C 6	21.8	87.2	27677	3	US-09-949-016-13055
C 7	21.8	87.2	84558	3	US-09-949-016-17572
C 8	21.8	87.2	174170	3	US-09-949-016-14810
C 9	21.8	87.2	174170	3	US-09-949-016-14811
C 10	21.8	87.2	174318	3	US-09-949-016-11880
C 11	21.8	87.2	174318	3	US-09-949-016-14812
C 12	21.8	87.2	174318	3	US-09-949-016-14813
C 13	20.8	83.2	24139	3	US-09-949-016-14675
C 14	20.8	83.2	125188	3	US-09-949-016-11980
C 15	20.2	80.8	138	3	US-09-513-9999C-21508
C 16	20.2	80.8	206	3	US-09-513-9999C-18230
C 17	20.2	80.8	601	3	US-09-949-016-17632
C 18	20.2	80.8	601	3	US-09-949-016-17633
C 19	20.2	80.8	601	3	US-09-949-016-17635
C 20	20.2	80.8	601	3	US-09-949-016-36195
C 21	20.2	80.8	601	3	US-09-949-016-36196
C 22	20.2	80.8	601	3	US-09-949-016-36197
C 23	20.2	80.8	601	3	US-09-949-016-41289
C 24	20.2	80.8	601	3	US-09-949-016-41290

25	20.2	80.8	601	3	US-09-949-016-48718	Sequence 48718, A
C 26	20.2	80.8	601	3	US-09-949-016-54603	Sequence 54603, A
C 27	20.2	80.8	601	3	US-09-949-016-54604	Sequence 54604, A
C 28	20.2	80.8	601	3	US-09-949-016-54605	Sequence 54605, A
C 29	20.2	80.8	601	3	US-09-949-016-54635	Sequence 54635, A
C 30	20.2	80.8	601	3	US-09-949-016-54636	Sequence 54636, A
C 31	20.2	80.8	601	3	US-09-949-016-54637	Sequence 54637, A
C 32	20.2	80.8	601	3	US-09-949-016-54667	Sequence 54667, A
C 33	20.2	80.8	601	3	US-09-949-016-54668	Sequence 54668, A
C 34	20.2	80.8	601	3	US-09-949-016-54669	Sequence 54669, A
C 35	20.2	80.8	601	3	US-09-949-016-56301	Sequence 56301, A
C 36	20.2	80.8	601	3	US-09-949-016-56302	Sequence 56302, A
C 37	20.2	80.8	601	3	US-09-949-016-59272	Sequence 59272, A
C 38	20.2	80.8	601	3	US-09-949-016-63344	Sequence 63344, A
C 39	20.2	80.8	601	3	US-09-949-016-119032	Sequence 119032, A
C 40	20.2	80.8	601	3	US-09-949-016-119033	Sequence 119033, A
C 41	20.2	80.8	601	3	US-09-949-016-119034	Sequence 119034, A
C 42	20.2	80.8	601	3	US-09-949-016-119068	Sequence 119068, A
C 43	20.2	80.8	601	3	US-09-949-016-119069	Sequence 119069, A
C 44	20.2	80.8	601	3	US-09-949-016-119070	Sequence 119070, A
C 45	20.2	80.8	601	3	US-09-949-016-119104	Sequence 119104, A
C 46	20.2	80.8	601	3	US-09-949-016-119105	Sequence 119105, A
C 47	20.2	80.8	601	3	US-09-949-016-119106	Sequence 119106, A
C 48	20.2	80.8	601	3	US-09-949-016-119140	Sequence 119140, A
C 49	20.2	80.8	601	3	US-09-949-016-119141	Sequence 119141, A
C 50	20.2	80.8	601	3	US-09-949-016-119142	Sequence 119142, A
C 51	20.2	80.8	601	3	US-09-949-016-119176	Sequence 119176, A
C 52	20.2	80.8	601	3	US-09-949-016-119177	Sequence 119177, A
C 53	20.2	80.8	601	3	US-09-949-016-119178	Sequence 119178, A
C 54	20.2	80.8	601	3	US-09-949-016-119212	Sequence 119212, A
C 55	20.2	80.8	601	3	US-09-949-016-119213	Sequence 119213, A
C 56	20.2	80.8	601	3	US-09-949-016-119214	Sequence 119214, A
C 57	20.2	80.8	601	3	US-09-949-016-135443	Sequence 135443, A
C 58	20.2	80.8	601	3	US-09-949-016-135444	Sequence 135444, A
C 59	20.2	80.8	601	3	US-09-949-016-175158	Sequence 175158, A
C 60	20.2	80.8	601	3	US-09-949-016-183999	Sequence 183999, A
C 61	20.2	80.8	601	3	US-09-949-016-184000	Sequence 184000, A
C 62	20.2	80.8	601	3	US-09-949-016-184104	Sequence 184104, A
C 63	20.2	80.8	601	3	US-09-949-016-184105	Sequence 184105, A
C 64	20.2	80.8	601	3	US-09-949-016-184209	Sequence 184209, A
C 65	20.2	80.8	601	3	US-09-949-016-184210	Sequence 184210, A
C 66	20.2	80.8	601	3	US-09-949-016-184314	Sequence 184314, A
C 67	20.2	80.8	601	3	US-09-949-016-184315	Sequence 184315, A
C 68	20.2	80.8	601	3	US-09-949-016-190842	Sequence 190842, A
C 69	20.2	80.8	601	3	US-09-949-002-1726	Sequence 1726, Ap
C 70	20.2	80.8	601	3	US-09-949-002-3436	Sequence 3436, Ap
C 71	20.2	80.8	601	3	US-09-949-002-3437	Sequence 3437, Ap
C 72	20.2	80.8	601	3	US-09-949-002-3801	Sequence 3801, Ap
C 73	20.2	80.8	601	3	US-09-949-002-6416	Sequence 6416, Ap
C 74	20.2	80.8	601	3	US-09-949-002-7692	Sequence 7692, Ap
C 75	20.2	80.8	601	3	US-09-949-002-9453	Sequence 9453, Ap
C 76	20.2	80.8	601	3	US-09-949-002-9454	Sequence 9454, Ap
C 77	20.2	80.8	1110	3	US-09-511-6258-43	Sequence 43, Appl
C 78	20.2	80.8	4460	3	US-09-103-875-4	Sequence 4, Appl1
C 79	20.2	80.8	5819	3	US-09-949-016-14050	Sequence 14050, A
C 80	20.2	80.8	7482	3	US-09-949-016-15564	Sequence 15564, A
C 81	20.2	80.8	10134	3	US-09-949-016-16668	Sequence 16668, A
C 82	20.2	80.8	10784	3	US-09-949-016-12913	Sequence 12913, A
C 83	20.2	80.8	12921	3	US-09-949-016-14446	Sequence 14446, A
C 84	20.2	80.8	15376	3	US-09-949-016-11937	Sequence 11937, A
C 85	20.2	80.8	15722	3	US-09-949-016-16709	Sequence 16709, A
C 86	20.2	80.8	16621	3	US-09-949-016-13708	Sequence 13708, A
C 87	20.2	80.8	17013	3	US-09-949-016-16229	Sequence 16229, A
C 88	20.2	80.8	17425	3	US-09-511-6258-5	Sequence 5, Appl1
C 89	20.2	80.8	20407	3	US-09-949-002-666	Sequence 666, App
C 90	20.2	80.8	20495	3	US-09-949-016-17198	Sequence 17198, A
C 91	20.2	80.8	23015	3	US-09-949-002-726	Sequence 726, App
C 92	20.2	80.8	23953	3	US-09-949-016-13500	Sequence 13500, A
C 93	20.2	80.8	24154	3	US-09-949-016-16374	Sequence 16374, A
C 94	20.2	80.8	24167	3	US-09-949-016-17542	Sequence 17542, A
C 95	20.2	80.8	27968	3	US-09-949-016-15191	Sequence 15191, A
C 96	20.2	80.8	27968	3	US-09-949-016-15192	Sequence 15192, A
C 97	20.2	80.8	29994	3	US-09-949-016-14836	Sequence 14836, A



```
Best Local Similarity 96.0%; Pred. No. 1.3;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    ||||| ||||| ||||| ||||| |||||
Db 69754 AAAAAAAAAATCAGCGCCAGGCATGG 69778

RESULT 4
US-09-949-016-13646
; Sequence 13646, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13646
; LENGTH: 23640
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(23640)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13646

Query Match 87.2%; Score 21.8; DB 3; Length 23640;
Best Local Similarity 92.0%; Pred. No. 5.3;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    ||||| ||||| ||||| ||||| |||||
Db 18856 AAAAAAAAAATCAGCGCCAGGCATGG 18880

RESULT 5
US-09-949-016-14682
; Sequence 14682, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14682
; LENGTH: 27592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(27592)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14682

Query Match 87.2%; Score 21.8; DB 3; Length 27592;
Best Local Similarity 92.0%; Pred. No. 5.4;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    ||||| ||||| ||||| ||||| |||||
Db 26968 AAAAAAAAAATCAGCGCCAGGCATGG 26992

RESULT 6
US-09-949-016-13055
; Sequence 13055, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13055
; LENGTH: 27767
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(27767)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13055

Query Match 87.2%; Score 21.8; DB 3; Length 27767;
Best Local Similarity 92.0%; Pred. No. 5.4;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
    ||||| ||||| ||||| ||||| |||||
Db 26968 AAAAAAAAAATCAGCGCCAGGCATGG 26992

RESULT 7
US-09-949-016-15752/c
; Sequence 15752, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15752
; LENGTH: 84558
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(84558)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15752
```

```
Query Match      87.2%; Score 21.8; DB 3; Length 84558;
Best Local Similarity 92.0%; Pred. No. 6.2;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 14022 AAAAAAAAAACAGCCAGGCATGG 13998

RESULT 8
US-09-949-016-14810
; Sequence 14810, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14810
; LENGTH: 174170
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(174170)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14810

Query Match      87.2%; Score 21.8; DB 3; Length 174170;
Best Local Similarity 92.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 14074 AAAAAAAAAATCAGGCCAGGCATGG 14098

RESULT 9
US-09-949-016-14811
; Sequence 14811, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14811
; LENGTH: 174170
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(174170)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14811

; LOCATION: (1)-(174170)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14811

Query Match      87.2%; Score 21.8; DB 3; Length 174318;
Best Local Similarity 92.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 14222 AAAAAAAAAATCAGGCCAGGCATGG 14246

RESULT 11
US-09-949-016-14812
; Sequence 14812, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14812
; LENGTH: 174318
; TYPE: DNA
```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(174318)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14812

Query Match      87.2%; Score 21.8; DB 3; Length 174318;
Best Local Similarity 92.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 14222 AAAAAAAAAATCAGGCCAGGCATGG 14246

RESULT 12
US-09-949-016-14813
; Sequence 14813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14813
; LENGTH: 174318
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(174318)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14813

Query Match      87.2%; Score 21.8; DB 3; Length 174318;
Best Local Similarity 92.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 14222 AAAAAAAAAATCAGGCCAGGCATGG 14246

RESULT 13
US-09-949-016-14675
; Sequence 14675, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14675
; LENGTH: 24139
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14675

Query Match      83.2%; Score 20.8; DB 3; Length 24139;
Best Local Similarity 91.7%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATG 24
   |||||
Db 14130 AAAAAAAAAATCAGGCCAGGCATG 14153

RESULT 14
US-09-949-016-11980
; Sequence 11980, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11980
; LENGTH: 125188
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(125188)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11980

Query Match      83.2%; Score 20.8; DB 3; Length 125188;
Best Local Similarity 91.7%; Pred. No. 18;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATG 24
   |||||
Db 93166 AAAAAAAAAATCAGGCCAGGCATG 93189

RESULT 15
US-09-513-999C-21508/c
; Sequence 21508, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21508
; LENGTH: 138
; TYPE: DNA
```





; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17633  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17633

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 68 AAAAAAAAAAGGCCAGGCATGG 44

## RESULT 20

US-09-949-016-36195/c  
; Sequence 36195, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 36195  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-36195

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 360 AAAAAAAAAAGGCCAGGCATGG 336

## RESULT 21

US-09-949-016-36196/c  
; Sequence 36196, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 36196  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-36196

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 137 AAAAAAAAAAGGCCAGGCATGG 113

## RESULT 22

US-09-949-016-36197/c  
; Sequence 36197, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 36197  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-36197

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 68 AAAAAAAAAAGGCCAGGCATGG 44

## RESULT 23

US-09-949-016-41289/c  
; Sequence 41289, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 41289  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-41289

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| |||||||  
Db 333 AAAAAAAAAATCAGCGCCAGGCATGG 309

RESULT 24  
US-09-949-016-41290/c  
; Sequence 41290, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 41290  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-41290

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| |||||||  
Db 298 AAAAAAAAAATCAGCGCCAGGCATGG 274

RESULT 25  
US-09-949-016-48718  
; Sequence 48718, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48718  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48718

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| |||||||  
Db 307 AAAAAAAAAAGGCCAGGCATGG 331

RESULT 26  
US-09-949-016-54603/c  
; Sequence 54603, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 54603  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-54603

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| |||||||  
Db 360 AAAAAAAAAAGGCCAGGCATGG 336

RESULT 27  
US-09-949-016-54604/c  
; Sequence 54604, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 54604  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-54604

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| |||||||  
Db 137 AAAAAAAAAAGGCCAGGCATGG 113

```
RESULT 28
US-09-949-016-54605/c
; Sequence 54605, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54605
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54605

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 68 AAAAAAAAAAAGGCCAGGCATGG 44

RESULT 29
US-09-949-016-54635/c
; Sequence 54635, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54635
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54635

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 360 AAAAAAAAAAAGGCCAGGCATGG 336

RESULT 30
US-09-949-016-54636/c
; Sequence 54636, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54636
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54636

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 68 AAAAAAAAAAAGGCCAGGCATGG 44

RESULT 31
US-09-949-016-54637/c
; Sequence 54637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54637
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54637

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 137 AAAAAAAAAAAGGCCAGGCATGG 113

RESULT 32
US-09-949-016-54667/c
; Sequence 54667, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54667
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54667

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 68 AAAAAAAAAAAGGCCAGGCATGG 44
```



; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-56302

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 594 AAAAAAAAAATCAGCGCCAGGCATGG 570

## RESULT 37

US-09-949-016-59272/c  
; Sequence 59272, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59272  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-59272

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 586 AAAAAAAAAATCAGCGCCAGGCATGG 562

## RESULT 38

US-09-949-016-63344/c  
; Sequence 63344, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63344  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-63344

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 415 AAAAAAAAAATCAGCGCCAGGCATGG 391

## RESULT 39

US-09-949-016-119032/c  
; Sequence 119032, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 119032  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-119032

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 360 AAAAAAAAAATCAGCGCCAGGCATGG 336

## RESULT 40

US-09-949-016-119033/c  
; Sequence 119033, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 119033  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-119033

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25

```
Db      137 AAAAAAAAAAAGGCCAGGCATGG 113
||||||| | |||||||||
RESULT 41
US-09-949-016-119034/c
; Sequence 119034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119034
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119034
Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCAGGCCAGGCATGG 25
||||||| | |||||||||
Db      68 AAAAAAAAAAAGGCCAGGCATGG 44
||||||| | |||||||||

RESULT 42
US-09-949-016-119069/c
; Sequence 119069, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119069
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119069
Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCAGGCCAGGCATGG 25
||||||| | |||||||||
Db      137 AAAAAAAAAAAGGCCAGGCATGG 113
||||||| | |||||||||

RESULT 43
US-09-949-016-119070/c
; Sequence 119070, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119070
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119070
Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCAGGCCAGGCATGG 25
||||||| | |||||||||
Db      68 AAAAAAAAAAAGGCCAGGCATGG 44
||||||| | |||||||||

RESULT 45
US-09-949-016-119104/c
; Sequence 119104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 119104  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-119104

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| | |||||  
Db 360 AAAAAAAAAAAGCCAGGCATGG 336

RESULT 46  
US-09-949-016-119105/c  
; Sequence 119105, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 119105  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-119105

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| | |||||  
Db 137 AAAAAAAAAAAGCCAGGCATGG 113

RESULT 47  
US-09-949-016-119106/c  
; Sequence 119106, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14

;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 119106  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-119106

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| | |||||  
Db 68 AAAAAAAAAAAGCCAGGCATGG 44

RESULT 48  
US-09-949-016-119140/c  
; Sequence 119140, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 119140  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-119140

Query Match 80.8%; Score 20.2; DB 3; Length 601;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||||| | |||||  
Db 360 AAAAAAAAAAAGCCAGGCATGG 336

RESULT 49  
US-09-949-016-119141/c  
; Sequence 119141, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498



```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119141
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119141

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||||| | ||||| |||||
Db 137 AAAAAAAAAAAGGCCAGGCATGG 113

RESULT 50
US-09-949-016-119142/c
; Sequence 119142, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119142
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119142

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   ||||||| | ||||| |||||
Db 68 AAAAAAAAAAAGGCCAGGCATGG 44

Search completed: February 3, 2006, 16:31:48
Job time : 86.6667 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-1

Perfect score: 25

Sequence: 1 aaaaaaaaaacagcgccagcgatg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	US-10-719-900-1
2	23.4	93.6	25	8	US-10-719-900-2
3	23.4	93.6	41309	7	US-10-741-601-5771
4	21.8	87.2	504	4	US-09-925-065A-287766
5	21.8	87.2	510	3	US-09-796-692-4519
6	21.8	87.2	510	3	US-09-796-692-4519
7	21.8	87.2	510	3	US-10-040-862-4519
8	21.8	87.2	510	5	US-10-040-862-4519
9	21.8	87.2	510	6	US-10-057-475B-4519
10	21.8	87.2	510	6	US-10-154-884B-4519
11	21.8	87.2	510	6	US-10-154-884B-4519
12	21.8	87.2	510	6	US-10-154-884B-4519
13	21.8	87.2	510	8	US-10-764-324-4519
14	21.8	87.2	510	8	US-10-764-324-4519
15	21.8	87.2	516	3	US-09-796-692-4518
16	21.8	87.2	516	5	US-10-040-862-4518
17	21.8	87.2	516	6	US-10-057-475B-4518
18	21.8	87.2	516	6	US-10-154-884B-4518
19	21.8	87.2	516	8	US-10-764-324-4518
20	21.8	87.2	590	4	US-09-925-065A-46634
21	21.8	87.2	591	4	US-09-925-065A-46634
22	21.8	87.2	591	5	US-10-027-632-277986
23	21.8	87.2	591	6	US-10-027-632-277986

24	21.8	87.2	715	5	US-10-027-632-20441	Sequence 20441, A	
25	21.8	87.2	715	6	US-10-027-632-20441	Sequence 20441, A	
C	26	21.8	1870	4	US-09-925-065A-37155	Sequence 37155, A	
27	21.8	87.2	2969	4	US-09-925-065A-28421	Sequence 28421, A	
28	21.8	87.2	2969	4	US-09-925-065A-60862	Sequence 60862, A	
C	29	21.8	3067	6	US-10-057-475B-10476	Sequence 10476, A	
C	30	21.8	3067	6	US-10-154-884B-10476	Sequence 10476, A	
C	31	21.8	3105	6	US-10-057-475B-10477	Sequence 10477, A	
C	32	21.8	3105	6	US-10-154-884B-10477	Sequence 10477, A	
C	33	21.8	3288	6	US-10-144-194A-37	Sequence 37, Appl	
C	34	21.8	3288	6	US-10-491-566-37	Sequence 37, Appl	
C	35	21.8	51657	6	US-10-057-475B-10475	Sequence 10475, A	
C	36	21.8	51657	6	US-10-154-884B-10475	Sequence 10475, A	
C	37	21.8	98642	5	US-10-087-192-1105	Sequence 1105, A	
C	38	21.8	98642	5	US-10-087-192-1864	Sequence 1864, A	
C	39	20.8	83.2	1315	4	US-09-925-065A-39390	Sequence 39390, A
C	40	20.8	83.2	25574	3	US-09-764-891-8307	Sequence 8307, A
41	20.8	83.2	25576	3	US-09-764-891-8306	Sequence 8306, A	
42	20.8	83.2	27893	6	US-10-017-161-757	Sequence 757, App	
43	20.8	83.2	401616	5	US-10-087-192-1630	Sequence 1630, A	
C	44	20.8	599001	9	US-10-317-869A-4	Sequence 4, Appl	
C	45	20.4	81.6	201	7	US-10-741-601-24054	Sequence 24054, A
C	46	20.4	81.6	201	8	US-10-741-600-64920	Sequence 64920, A
47	20.4	81.6	528	4	US-09-925-065A-163238	Sequence 163238, A	
C	48	20.4	81.6	546	5	US-10-027-632-201503	Sequence 201503, A
C	49	20.4	81.6	546	6	US-10-027-632-201503	Sequence 201503, A
C	50	20.4	81.6	573	4	US-09-925-065A-51348	Sequence 51348, A
C	51	20.4	81.6	590	4	US-09-925-065A-323966	Sequence 323966, A
52	20.4	81.6	1511	4	US-09-925-065A-681763	Sequence 681763, A	
C	53	20.4	15804	7	US-10-741-601-5663	Sequence 5663, A	
C	54	20.4	15804	8	US-10-741-600-17690	Sequence 17690, A	
C	55	20.4	24446	7	US-10-741-601-5740	Sequence 5740, A	
C	56	20.4	24446	8	US-10-741-600-17889	Sequence 17889, A	
C	57	20.4	41907	3	US-09-967-013-5	Sequence 5, Appl	
58	20.4	81.6	210204	8	US-10-723-860-1746	Sequence 1746, A	
59	20.4	81.6	210204	9	US-10-756-149-1685	Sequence 1685, A	
C	60	20.2	80.8	201	8	US-10-741-600-34716	Sequence 34716, A
61	20.2	80.8	344	3	US-09-803-719-718	Sequence 718, App	
62	20.2	80.8	344	9	US-10-779-543-13988	Sequence 13988, A	
C	63	20.2	80.8	384	3	US-09-995-793A-14	Sequence 14, Appl
64	20.2	80.8	440	4	US-09-925-065A-917840	Sequence 917840, A	
65	20.2	80.8	440	4	US-09-925-065A-944841	Sequence 944841, A	
66	20.2	80.8	463	8	US-10-674-1241-18846	Sequence 18846, A	
C	67	20.2	80.8	465	3	US-09-867-701-4978	Sequence 4978, A
68	20.2	80.8	471	4	US-09-925-065A-232436	Sequence 232436, A	
69	20.2	80.8	471	4	US-09-925-065A-232437	Sequence 232437, A	
70	20.2	80.8	471	4	US-09-925-065A-232438	Sequence 232438, A	
71	20.2	80.8	480	9	US-10-779-543-17707	Sequence 17707, A	
72	20.2	80.8	484	3	US-09-864-761-11717	Sequence 11717, A	
73	20.2	80.8	514	4	US-09-925-065A-599935	Sequence 599935, A	
74	20.2	80.8	514	4	US-09-925-065A-599936	Sequence 599936, A	
75	20.2	80.8	516	5	US-10-027-632-289481	Sequence 289481, A	
76	20.2	80.8	516	6	US-10-027-632-289481	Sequence 289481, A	
77	20.2	80.8	526	9	US-10-027-632-20631	Sequence 20631, A	
78	20.2	80.8	530	4	US-09-925-065A-939463	Sequence 939463, A	
79	20.2	80.8	530	4	US-09-925-065A-939464	Sequence 939464, A	
80	20.2	80.8	537	5	US-10-027-632-78935	Sequence 78935, A	
81	20.2	80.8	537	5	US-10-027-632-78935	Sequence 78935, A	
82	20.2	80.8	537	5	US-10-027-632-306937	Sequence 306937, A	
83	20.2	80.8	537	5	US-10-027-632-306938	Sequence 306938, A	
84	20.2	80.8	537	6	US-10-027-632-78934	Sequence 78934, A	
85	20.2	80.8	537	6	US-10-027-632-78935	Sequence 78935, A	
86	20.2	80.8	537	6	US-10-027-632-306937	Sequence 306937, A	
87	20.2	80.8	537	6	US-10-027-632-306938	Sequence 306938, A	
88	20.2	80.8	543	4	US-09-925-065A-17381	Sequence 17381, A	
C	89	20.2	80.8	556	4	US-09-925-065A-906458	Sequence 906458, A
90	20.2	80.8	564	4	US-09-925-065A-580000	Sequence 580000, A	
C	91	20.2	80.8	570	5	US-10-027-632-24416	Sequence 24416, A
C	92	20.2	80.8	570	6	US-10-027-632-24416	Sequence 24416, A
C	93	20.2	80.8	573	5	US-10-027-632-277986	Sequence 277986, A
C	94	20.2	80.8	573	6	US-10-027-632-277986	Sequence 277986, A
C	95	20.2	80.8	599	4	US-09-925-065A-662607	Sequence 662607, A
C	96	20.2	80.8	606	4	US-09-925-065A-770354	Sequence 770354, A

97	20.2	80.8	609	4	US-09-925-065A-105459
98	20.2	80.8	609	4	US-09-925-065A-105460
c 99	20.2	80.8	630	4	US-09-925-065A-937990
100	20.2	80.8	630	5	US-10-027-632-129451
101	20.2	80.8	630	6	US-10-027-632-129451
102	20.2	80.8	631	4	US-09-935-065A-940220
103	20.2	80.8	639	4	US-09-925-065A-923002
c 104	20.2	80.8	640	4	US-09-925-065A-53378
c 105	20.2	80.8	645	4	US-09-925-065A-732668
c 106	20.2	80.8	645	4	US-09-925-065A-732669
107	20.2	80.8	703	4	US-09-935-065A-940164
108	20.2	80.8	703	4	US-09-935-065A-940165
c 109	20.2	80.8	703	4	US-09-925-065A-955134
110	20.2	80.8	728	4	US-09-925-065A-215454
111	20.2	80.8	728	4	US-09-925-065A-215454
c 112	20.2	80.8	771	5	US-10-027-632-150013
c 113	20.2	80.8	771	6	US-10-027-632-150013
c 114	20.2	80.8	789	4	US-09-935-065A-953148
c 115	20.2	80.8	795	5	US-10-027-632-164373
c 116	20.2	80.8	795	6	US-10-027-632-164373
117	20.2	80.8	841	5	US-10-027-632-161160
118	20.2	80.8	841	6	US-10-027-632-161160
119	20.2	80.8	855	5	US-10-027-632-29404
c 120	20.2	80.8	855	6	US-10-027-632-29404
c 121	20.2	80.8	856	5	US-10-027-632-162444
c 122	20.2	80.8	856	5	US-10-027-632-162445
c 123	20.2	80.8	856	6	US-10-027-632-162444
c 124	20.2	80.8	856	6	US-10-027-632-162445
125	20.2	80.8	957	5	US-10-027-632-258749
126	20.2	80.8	957	5	US-10-027-632-258750
127	20.2	80.8	957	6	US-10-027-632-258750
128	20.2	80.8	957	6	US-10-027-632-258750
c 129	20.2	80.8	974	4	US-09-925-065A-54818
c 130	20.2	80.8	1012	4	US-09-935-065A-17445
c 131	20.2	80.8	1054	4	US-09-935-065A-297766
c 132	20.2	80.8	1103	5	US-10-027-632-30415
c 133	20.2	80.8	1103	6	US-10-027-632-30415
c 134	20.2	80.8	1131	3	US-09-764-878-260
c 135	20.2	80.8	1131	3	US-09-764-878-260
c 136	20.2	80.8	1131	3	US-09-764-878-260
c 137	20.2	80.8	1131	5	US-10-079-854-260
c 138	20.2	80.8	1131	6	US-10-079-854-260
c 139	20.2	80.8	1140	5	US-10-027-632-254873
140	20.2	80.8	1140	6	US-10-027-632-254873
141	20.2	80.8	1264	5	US-10-001-857-59
142	20.2	80.8	1507	4	US-09-935-065A-700674
143	20.2	80.8	1512	4	US-09-935-065A-680429
144	20.2	80.8	1512	4	US-09-935-065A-680430
145	20.2	80.8	1512	4	US-09-935-065A-680431
146	20.2	80.8	1512	4	US-09-935-065A-680433
147	20.2	80.8	1981	4	US-09-925-065A-717860
148	20.2	80.8	1981	4	US-09-935-065A-717861
c 149	20.2	80.8	2021	4	US-09-935-065A-27392
c 150	20.2	80.8	6186	3	US-09-764-860-1075
c 151	20.2	80.8	6186	5	US-10-074-095-1075

## ALIGNMENTS

```

RESULT 1
US-10-719-900-1
; Sequence 1, Application US/10719900
; Publication NO. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V

```

Sequence 105459, Sequence 105460, Sequence 937990, Sequence 129451, Sequence 129451, Sequence 940220, Sequence 940220, Sequence 523002, Sequence 53378, A Sequence 733668, Sequence 733669, Sequence 940164, Sequence 940165, Sequence 955134, Sequence 21545, A Sequence 21546, A Sequence 150013, Sequence 150013, Sequence 950148, Sequence 164373, Sequence 164373, Sequence 162444, Sequence 162444, Sequence 162445, Sequence 162445, Sequence 258749, Sequence 258750, Sequence 258749, Sequence 258750, Sequence 5818, A Sequence 17445, A Sequence 297786, A Sequence 30415, A Sequence 30415, A Sequence 260, App Sequence 260, App Sequence 939, App Sequence 939, App Sequence 254873, Sequence 254873, Sequence 59, App Sequence 59, App Sequence 706674, A Sequence 680429, Sequence 680430, Sequence 680431, Sequence 717860, Sequence 717861, Sequence 27392, A Sequence 1075, App Sequence 1075, App

```

; SEQ ID NO 1
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-1

Query Match          100.0%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

Qy 1 AAAAAAAAAATCAGGCGCAGGCATGG 25
    |||||
Db 1 AAAAAAAAAATCAGGCGCAGGCATGG 25

RESULT 2
US-10-719-900-2
; Sequence 2, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: xue Mei
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 2
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-2

Query Match          93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.74; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 1;

Qy 1 AAAAAAAAAATCAGGCGCAGGCATGG 25
    |||||
Db 1 AAAAAAAAAATCAGGCGCAGGCATGG 25

RESULT 3
US-10-741-601-5771/c
; Sequence 5771, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 5771
; LENGTH: 41309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5771

```

US-09-925-065A-287766  
; Sequence 287766, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 287766  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-287766

Query Match 87.2%; Score 21.8; DB 4; Length 504;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
DB 329 AAAAAAAAAATTAAGCGCCAGGCATGG 353

RESULT 5  
US-09-796-692-4519  
; Sequence 4519, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 9149  
; LENGTH: 510  
; TYPE: DNA

; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 4519  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (28)  
; OTHER INFORMATION: n=A,T,C or G  
; NAME/KEY: unsure  
; LOCATION: (242)  
; OTHER INFORMATION: n=A,T,C or G  
; NAME/KEY: unsure  
; LOCATION: (344)  
; OTHER INFORMATION: n=A,T,C or G  
; NAME/KEY: unsure  
; LOCATION: (472)  
; OTHER INFORMATION: n=A,T,C or G  
US-09-796-692-4519

Query Match 87.2%; Score 21.8; DB 3; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
DB 405 AAAAAAAAAATCAGCGCCAGGCATGG 429

RESULT 6  
US-09-796-692-9149  
; Sequence 9149, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 9149  
; LENGTH: 510  
; TYPE: DNA



OTHER INFORMATION: n=A,T,C or G  
FEATURE: unsure  
NAME/KEY: (481)  
LOCATION: (481)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE: unsure  
NAME/KEY: (504)  
LOCATION: (504)  
OTHER INFORMATION: n=A,T,C or G  
US-10-040-862-9149

Query Match 87.2%; Score 21.8; DB 5; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 405 AAAAAATAAATCAGCGCCAGGCACGG 429

RESULT 9  
US-10-057-475B-4519  
; Sequence 4519, Application US/10057475B  
; Publication No. US2004002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordonez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4519  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(510)  
; OTHER INFORMATION: n = G, A, C or T

Query Match 87.2%; Score 21.8; DB 6; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 405 AAAAAATAAATCAGCGCCAGGCACGG 429

RESULT 10  
US-10-057-475B-9149  
; Sequence 9149, Application US/10057475B  
; Publication No. US2004002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordonez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9149  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(510)  
; OTHER INFORMATION: n = G, A, C or T

Query Match 87.2%; Score 21.8; DB 6; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 405 AAAAAATAAATCAGCGCCAGGCACGG 429

RESULT 11  
US-10-154-884B-4519  
; Sequence 4519, Application US/10154884B  
; Publication No. US20040005561A1

```

; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4519
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: n = g, a, c or t
US-10-154-884B-4519

Query Match      87.2%; Score 21.8; DB 6; Length 510;
Best Local Similarity 92.08; Pred.No.6.5;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCAGGCCGCGCATGG 25
        ||||| ||||| ||||| ||||| |||||
Db     405 AAAAAATAAATCAGGCCGCGCATGG 429

RESULT 12
US-10-154-884B-9149
; Sequence 9149, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545

```



NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4519  
LENGTH: 510  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (28)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (242)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (344)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (472)  
OTHER INFORMATION: n=A,T,C or G  
US-10-764-324-4519

Query Match 87.2%; Score 21.8; DB 8; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCAGGATGG 25  
|||||  
Db 405 AAAAAATAATCAGCGCAGGACGG 429

RESULT 14

US-10-764-324-9149  
Sequence 9149, Application US/10764324  
Publication No. US20040175739A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
TITLE OF INVENTION: Hematological Malignancies  
FILE REFERENCE: 014058-013520US  
CURRENT APPLICATION NUMBER: US/10/764,324  
CURRENT FILING DATE: 2004-01-23  
PRIOR APPLICATION NUMBER: US/10/040,862  
PRIOR FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9149  
LENGTH: 510  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9149  
LENGTH: 510

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (472)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (481)  
OTHER INFORMATION: n=A,T,C or G  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (504)  
OTHER INFORMATION: n=A,T,C or G  
US-10-764-324-9149

Query Match 87.2%; Score 21.8; DB 8; Length 510;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCAGGATGG 25  
|||||  
Db 405 AAAAAATAATCAGCGCAGGACGG 429

RESULT 15

US-09-796-692-4518  
Sequence 4518, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4518  
LENGTH: 516  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-4518

Query Match 87.2%; Score 21.8; DB 3; Length 516;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCGATGG 25  
||||| ||||||| ||||||| |||  
Db 405 AAAAAATAATCAGCGCCAGGCGCG 429

RESULT 16  
US-10-040-862-4518  
; Sequence 4518, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4518  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-4518

Query Match 87.2%; Score 21.8; DB 5; Length 516;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCGATGG 25  
||||| ||||||| ||||||| |||  
Db 405 AAAAAATAATCAGCGCCAGGCGCG 429

RESULT 17  
US-10-057-475B-4518  
; Sequence 4518, Application US/10057475B  
; Publication No. US2004002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David

; APPLICANT: Wang, Aijun  
; APPLICANT: Ordenez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4518  
; LENGTH: 516  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-4518

Query Match 87.2%; Score 21.8; DB 6; Length 516;  
Best Local Similarity 92.0%; Pred. No. 6.5;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCGATGG 25  
||||| ||||||| ||||||| |||  
Db 405 AAAAAATAATCAGCGCCAGGCGCG 429

RESULT 18  
US-10-154-884B-4518  
; Sequence 4518, Application US/10154884B  
; Publication No. US2004005561A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013521US  
; CURRENT APPLICATION NUMBER: US/10/154,884B  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28

```
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4518
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-4518

Query Match      87.2%; Score 21.8; DB 6; Length 516;
Best Local Similarity 92.0%; Pred. No. 6.5;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
DB      405 AAAAAATTAATCAGCGCCAGGCATGG 429

RESULT 19
US-10-764-324-4518
; Sequence 4518, Application US/10764324
; Publication No. US2004017539A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/764,324
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4518
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-324-4518

Query Match      87.2%; Score 21.8; DB 8; Length 516;
```

```
Best Local Similarity 92.0%; Pred. No. 6.5;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
DB      405 AAAAAATTAATCAGCGCCAGGCATGG 429

RESULT 20
US-09-925-065A-46634/c
; Sequence 46634, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46634
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-46634

Query Match      87.2%; Score 21.8; DB 4; Length 590;
Best Local Similarity 92.0%; Pred. No. 6.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
DB      129 AAAAAAAATCTTTGCCAGGCATGG 105

RESULT 21
US-09-925-065A-677298
; Sequence 677298, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 677298
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
```

US-09-925-065A-677298

Query Match 87.2%; Score 21.8; DB 4; Length 591;  
Best Local Similarity 92.0%; Pred. No. 6.6;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||| ||| ||||| ||||| |||||  
Db 456 AAAAAAAAAATCAGCGCCAGGCATGG 480

RESULT 22

US-10-027-632-269686  
; Sequence 269686, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 269686

; LENGTH: 591

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-269686

Query Match 87.2%; Score 21.8; DB 5; Length 591;  
Best Local Similarity 92.0%; Pred. No. 6.6;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||| ||| ||||| ||||| |||||  
Db 64 AAAAAAAAAATCAGCGCCAGGCATGG 88

RESULT 23

US-10-027-632-269686  
; Sequence 269686, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 269686  
; LENGTH: 591  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-269686

Query Match 87.2%; Score 21.8; DB 6; Length 591;  
Best Local Similarity 92.0%; Pred. No. 6.6;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||| ||| ||||| ||||| |||||  
Db 64 AAAAAAAAAATCAGCGCCAGGCATGG 88

RESULT 24

US-10-027-632-20441  
; Sequence 20441, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 20441

; LENGTH: 715

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-20441

Query Match 87.2%; Score 21.8; DB 5; Length 715;  
Best Local Similarity 92.0%; Pred. No. 6.8;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
||||| ||| ||||| ||||| |||||  
Db 402 AAAAAAAAAATCAGCGCCAGGCATGG 426

RESULT 25

US-10-027-632-20441  
; Sequence 20441, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129

```

, , CURRENT APPLICATION NUMBER: US/10/027,632
, , CURRENT FILING DATE: 2002-04-30
, , PRIOR APPLICATION NUMBER: US 60/218,006
, , PRIOR FILING DATE: 2000-07-12
, , PRIOR APPLICATION NUMBER: US 60/198,676
, , PRIOR FILING DATE: 2000-04-20
, , PRIOR APPLICATION NUMBER: US 60/193,483
, , PRIOR FILING DATE: 2000-03-29
, , PRIOR APPLICATION NUMBER: US 60/185,218
, , PRIOR FILING DATE: 2000-02-24
, , PRIOR APPLICATION NUMBER: US 60/167,363
, , PRIOR FILING DATE: 1999-11-23
, , PRIOR APPLICATION NUMBER: US 60/156,358
, , PRIOR FILING DATE: 1999-09-28
, , PRIOR APPLICATION NUMBER: US 60/146,002
, , PRIOR FILING DATE: 1999-08-09
, , NUMBER OF SEQ ID NOS: 325720
, , SOFTWARE: FastSeq for Windows Version 4.0
, , SEQ ID NO 20441
, , LENGTH: 715
, , TYPE: DNA
, , ORGANISM: Human
US-10-027-632-20441

```

Query Match 87.2%; Score 21.8; DB 6; Length 715;  
Best Local Similarity 92.0%; Pred. No. 6.8;  
Matches 23: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCACGGCCAGGCATGG 25  
|||  
402 AAAAAATAATCACGGCCAGGCACGG 426  
Db

RESULT 26

```

US-09-925-065A-37155/c
; Sequence 37155, Application US/09925065A
; Publication NO. US20050228172A9
GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and
; TITLE OF INVENTION: Nucleotide Polymor
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,0
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: Fast-SEQ for Windows Version 4
; SEQ ID NO 37155
; LENGTH: 1870
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-37155

```

Query Match 87.2%; Score 21.8; DB 4; Length 1870;  
Best Local Similarity 92.0%; Pred. No. 7.9;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
|||  
Db 710 AAAAAAAAAATCAGGCCAGGCACGG 686

RESULT 27  
US-09-925-065A-28421

```

; Sequence 28421, Application US/09925065A
; Publication No. US20050228172A9
;
; GENERAL INFORMATION:
;
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FAST-SEQ for Windows Version 4.0
; SEQ ID NO 28421
; LENGTH: 2369
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-28421

```

Query Match	87.2%	Score 21.8;	DB 4;	Length 2969;
Best Local Similarity	92.0%	Pred. No. 8.5;		
Matches 23: Conservative		0; Mismatches 2;	Indels 0;	Gaps 0;

Qy 1 AAAAAAAAAATCACGGCCAGGCATGG 25  
|||  
1239 AAAAAAAAAATCATGGCCAGGCGTGG 1263  
pb

RESULT 28

```

US-09-925-065A-60862
; Sequence 60862, Application US/09925065A
; Publication No. US20050228172A9
;
GENERAL INFORMATION:
;
; APPLICANT: Wang, David G.
;
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
;
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 60862
; LENGTH: 2969
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-60862

```

Query Match	87.2%	Score 21.8;	DB 4;	Length 2969;
Best Local Similarity	92.0%	Pred. No. 8.5;		
Matches 23: Conservative	0;	Mismatches	2;	Indels 0;
Gaps	0;			

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
|||  
pb 1239 AAAAAAAAAATCATGGCCAGGCGTGG 1263

```
RESULT 29
US-10-057-475B-10476/c
; Sequence 10476, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US/10/057,475B
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10476
; LENGTH: 3067
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-10476

Query Match      87.2%; Score 21.8; DB 6; Length 3067;
Best Local Similarity 92.0%; Pred. No. 8.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATGG 25
DB 1184 AAAAAATAATCAGCGCCAGCGCATGG 1160

RESULT 30
US-10-154-884B-10476/c
; Sequence 10476, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT FILING DATE: 2002-05-23
; CURRENT APPLICATION NUMBER: US/10/154,884B
```

```
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10476
; LENGTH: 3067
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-10476

Query Match      87.2%; Score 21.8; DB 6; Length 3067;
Best Local Similarity 92.0%; Pred. No. 8.6;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGCGCATGG 25
DB 1184 AAAAAATAATCAGCGCCAGCGCATGG 1160

RESULT 31
US-10-057-475B-10477/c
; Sequence 10477, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US/10/057,475B
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
```

; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10477  
; LENGTH: 3105  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-10477

Query Match 87.2%; Score 21.8; DB 6; Length 3105;  
Best Local Similarity 92.0%; Pred. No. 8.6;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 1177 AAAAAATAATCAGCGCCAGGCATGG 1153

## RESULT 32

US-10-154-884B-10477/c  
; Sequence 10477, Application US/10154884B  
; Publication No. US20040005561A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013521US  
; CURRENT APPLICATION NUMBER: US/10/154,884B  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 11290  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10477  
; LENGTH: 3105  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-154-884B-10477

Query Match 87.2%; Score 21.8; DB 6; Length 3105;  
Best Local Similarity 92.0%; Pred. No. 8.6;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 1177 AAAAAATAATCAGCGCCAGGCATGG 1153

## RESULT 33

US-10-144-194A-37  
; Sequence 37, Application US/10144194A  
; Publication No. US20030215809A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies Inc  
; TITLE OF INVENTION: Regulated Breast Cancer Genes  
; FILE REFERENCE: 3U 103 R1  
; CURRENT APPLICATION NUMBER: US/10/144,194A  
; CURRENT FILING DATE: 2002-06-12  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 37  
; LENGTH: 3288  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (267)..(653)  
US-10-144-194A-37

Query Match 87.2%; Score 21.8; DB 6; Length 3288;  
Best Local Similarity 92.0%; Pred. No. 8.7;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 2979 AAAAAATAATCTTGGCCAGGCATGG 3003

## RESULT 34

US-10-491-566-37  
; Sequence 37, Application US/10491566  
; Publication No. US20040249144A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies Inc  
; TITLE OF INVENTION: Regulated Breast Cancer Genes  
; FILE REFERENCE: 3U 103 R1  
; CURRENT APPLICATION NUMBER: US/10/491,566  
; CURRENT FILING DATE: 2004-04-05  
; NUMBER OF SEQ ID NOS: 148  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 37  
; LENGTH: 3288  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (267)..(653)  
; OTHER INFORMATION:  
US-10-491-566-37

Query Match 87.2%; Score 21.8; DB 8; Length 3288;  
Best Local Similarity 92.0%; Pred. No. 8.7;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
|||||  
Db 2979 AAAAAATAATCTTGGCCAGGCATGG 3003

## RESULT 35

US-10-057-475B-10475/c  
; Sequence 10475, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun



```

; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10475
; LENGTH: 51657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(51657)
; OTHER INFORMATION: n = g, a, c or t
US-10-057-475B-10475

Query Match      87.2%; Score 21.8; DB 6; Length 51657;
Best Local Similarity 92.0%; Pred. No. 13;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAATCAGCGCCAGGCATCG 25
        ||||| ||||| ||||| ||||| |||||
Db      28204 AAAAAATAATCAGCGCCAGGCATCG 28180

RESULT 36
US-10-154-884B-10475/c
; Sequence 10475, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303

```

```

US-10-087-192-1864/C
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8307
; LENGTH: 25574
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8307
Query Match      83.2%; Score 20.8; DB 3; Length 25574;
Best Local Similarity 91.7%; Pred. No. 34;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   |||||
Db 12227 AAAAAAAAAATCAAGGCCAGGCCTG 12250

RESULT 41
US-09-764-891-8306
; Sequence 8306, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8306
; LENGTH: 25576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8306
Query Match      83.2%; Score 20.8; DB 3; Length 25576;
Best Local Similarity 91.7%; Pred. No. 34;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   |||||
Db 12225 AAAAAAAAAATCAAGGCCAGGCCTG 12248

RESULT 42
US-10-017-161-757
; Sequence 757, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 757
; LENGTH: 27893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-39390
Query Match      83.2%; Score 20.8; DB 4; Length 1315;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 660 AAATAAATCAGCGCCAGGCATGG 637

RESULT 40
US-09-764-891-8307
; Sequence 8307, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1864
; LENGTH: 350764
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1864
Query Match      87.2%; Score 21.8; DB 5; Length 350764;
Best Local Similarity 92.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 23134 AAAAAAAAAAAGCGCCAGGCATGG 23110

RESULT 39
US-09-925-065A-39390/C
; Sequence 39390, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39390
; LENGTH: 1315
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-39390
Query Match      83.2%; Score 20.8; DB 4; Length 1315;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 660 AAATAAATCAGCGCCAGGCATGG 637

RESULT 40
US-09-764-891-8307
; Sequence 8307, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8306
; LENGTH: 25576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8306
Query Match      83.2%; Score 20.8; DB 3; Length 25576;
Best Local Similarity 91.7%; Pred. No. 34;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   |||||
Db 12225 AAAAAAAAAATCAAGGCCAGGCCTG 12248

RESULT 42
US-10-017-161-757
; Sequence 757, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 757
; LENGTH: 27893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-39390
Query Match      83.2%; Score 20.8; DB 4; Length 1315;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 660 AAATAAATCAGCGCCAGGCATGG 637

RESULT 40
US-09-764-891-8307
; Sequence 8307, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1864
; LENGTH: 350764
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1864
Query Match      87.2%; Score 21.8; DB 5; Length 350764;
Best Local Similarity 92.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 23134 AAAAAAAAAAAGCGCCAGGCATGG 23110

RESULT 39
US-09-925-065A-39390/C
; Sequence 39390, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39390
; LENGTH: 1315
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-39390
Query Match      83.2%; Score 20.8; DB 4; Length 1315;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25
   |||||
Db 660 AAATAAATCAGCGCCAGGCATGG 637

RESULT 40
US-09-764-891-8307
; Sequence 8307, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8306
; LENGTH: 25576
; TYPE: DNA
; ORGANISM: Homo
```

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: source  
LOCATION: (1)..(27893)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)..(252)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11364)..(11463)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (14312)..(14361)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18284)..(18497)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18620)..(19553)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (27613)..(27693)  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (3166)..(3265)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (3472)  
FEATURE:  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (8025)..(8124)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (9388)  
FEATURE:  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (9391)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (9985)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (13824)..(13923)  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (24285)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (24410)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (24443)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (24787)..(24886)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (27429)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (27432)

OTHER INFORMATION: a, t, c, g, unknown or other  
US-10-017-161-757  
Query Match 83.2%; Score 20.8; DB 6; Length 27893;  
Best Local Similarity 91.7%; Pred. No. 35;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 27133 AAAAAAAAAATCAGCGCCAGGCATGG 27156  
RESULT 43  
US-10-087-192-1630/c  
Sequence 1630, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: CANCER  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1630  
LENGTH: 401616  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(401616)  
OTHER INFORMATION: n = A, T, C or G  
US-10-087-192-1630  
Query Match 83.2%; Score 20.8; DB 5; Length 401616;  
Best Local Similarity 91.7%; Pred. No. 53;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 140060 AAAAAAAAAATCAGCGCCAGGCATGG 140037  
RESULT 44  
US-10-317-869A-4/c  
Sequence 4, Application US/10317869A  
Publication No. US20050101000A1  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
APPLICANT: Kenneth W. Dobie  
TITLE OF INVENTION: MODULATION OF PHOSPHODIESTERASE 4B EXPRESSION  
FILE REFERENCE: RTS-0429  
CURRENT APPLICATION NUMBER: US/10/317,869A  
CURRENT FILING DATE: 2002-12-11  
NUMBER OF SEQ ID NOS: 113  
SEQ ID NO 4  
LENGTH: 599001  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 55607-55706  
LOCATION: 360227-360326, 363103-363202, 363428, 366093-366192,  
LOCATION: 368528, 447532, 452653-452752, 457192, 457207-457306, 461808-461907,  
LOCATION: 461921, 461928, 461942  
OTHER INFORMATION: n = A, T, C or G  
US-10-317-869A-4

```
Query Match      83.2%; Score 20.8; DB 9; Length 599001;
Best Local Similarity 91.7%; Pred. No. 56;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCAGCCGCGCATGG 25
   |||||
Db 330645 AAAAAAAAAATCAGCCGCGCATGG 330622

RESULT 45
US-10-741-601-24054/c
; Sequence 24054, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24054
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-24054

Query Match      81.6%; Score 20.4; DB 7; Length 201;
Best Local Similarity 95.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCGCGCAGCA 22
   |||||
Db 109 AAAAAAAAAATCAGCCGCGCAGCA 88

RESULT 46
US-10-741-600-64920/c
; Sequence 64920, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64920
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-64920

Query Match      81.6%; Score 20.4; DB 8; Length 201;
Best Local Similarity 95.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCGCGCAGCA 22
   |||||
Db 109 AAAAAAAAAATCAGCCGCGCAGCA 88

RESULT 47
US-09-925-065A-163238
; Sequence 163238, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 163238
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-163238

Query Match      81.6%; Score 20.4; DB 4; Length 528;
Best Local Similarity 95.5%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCGCGCAGCA 22
   |||||
Db 280 AAAAAAAAAATCAGCCGCGCAGCA 301

RESULT 48
US-10-027-632-201503/c
; Sequence 201503, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201503
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-201503

Query Match      81.6%; Score 20.4; DB 5; Length 546;
Best Local Similarity 95.5%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCCGCGCAGCA 22
   |||||
Db 23 AAAAAAAAAATCAGCCGCGCAGCA 2

RESULT 49
```

```
US-10-027-632-201503/c
; Sequence 201503, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201503
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-201503
```

```
Query Match      81.6%; Score 20.4; DB 6; Length 546;
Best Local Similarity 95.5%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCAGGCCAGGCA 22
   |||||
Db 360 AAAAAAAAAATCAGGCCAGGCA 381

Search completed: February 3, 2006, 15:43:18
Job time : 370.556 secs
```

```
US-10-027-632-201503
; Sequence 201503, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51348
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(573)
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-51348
```

```
RESULT 50
US-09-925-065A-51348
; Sequence 51348, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51348
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(573)
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-51348
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-1

Perfect score: 25

Sequence: 1 aaaaaaaatcacggccagcatgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA New:  
1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.4	93.6	201	7	US-10-995-561-79948
C 2	23.4	93.6	41309	7	US-10-995-561-13487
C 3	20.4	81.6	201	7	US-10-995-561-72586
C 4	20.4	81.6	201	7	US-10-995-561-84366
C 5	20.4	81.6	15804	7	US-10-995-561-13294
C 6	20.4	81.6	24446	7	US-10-995-561-13436
C 7	20.4	81.6	40000	7	US-10-995-561-13513
C 8	20.4	81.6	153142	8	US-11-121-086-27
C 9	20.2	80.8	201	7	US-10-995-561-33166
C 10	20.2	80.8	201	7	US-10-995-561-50016
C 11	20.2	80.8	32157	7	US-10-995-561-13352
C 12	20.2	80.8	44229	8	US-11-124-368A-2910
C 13	20.2	80.8	60844	7	US-10-995-561-13359
C 14	20.2	80.8	66280	8	US-11-124-368A-2911
C 15	20.2	80.8	86950	7	US-10-857-780-5
C 16	20.2	80.8	101046	7	US-10-995-561-13330
C 17	20.2	80.8	128978	7	US-10-775-169-345
C 18	20.2	80.8	141121	7	US-10-995-561-13262
C 19	20.2	80.8	160226	8	US-11-121-086-29
C 20	20.2	80.8	160226	8	US-11-121-086-29
C 21	20.2	80.8	168656	8	US-11-112-908-59
C 22	20.2	80.8	170285	8	US-11-112-908-58

C 23	20.2	80.8	179666	8	US-11-121-086-67	Sequence 67, Appl
C 24	20.2	80.8	179892	8	US-11-112-908-39	Sequence 39, Appl
C 25	20.2	80.8	415117	7	US-10-995-561-13274	Sequence 13274, A
C 26	20.2	80.8	415117	7	US-10-995-561-13274	Sequence 13274, A
C 27	19.8	79.2	201	7	US-10-995-561-36951	Sequence 36951, A
C 28	19.8	79.2	201	7	US-10-995-561-50172	Sequence 50172, A
C 29	19.8	79.2	170995	8	US-11-121-086-35	Sequence 35, Appl
C 30	19.8	79.2	201990	7	US-10-995-561-13303	Sequence 13303, A
C 31	19.2	76.8	25	8	US-11-121-849-249080	Sequence 249080, A
C 32	19.2	76.8	25	8	US-11-121-849-249081	Sequence 249081, A
C 33	19.2	76.8	25	8	US-11-121-849-250266	Sequence 250266, A
C 34	19.2	76.8	25	8	US-11-121-849-250267	Sequence 250267, A
C 35	19.2	76.8	201	7	US-10-995-561-28311	Sequence 28311, A
C 36	19.2	76.8	201	7	US-10-995-561-28314	Sequence 28314, A
C 37	19.2	76.8	201	7	US-10-995-561-28356	Sequence 28356, A
C 38	19.2	76.8	201	7	US-10-995-561-32893	Sequence 32893, A
C 39	19.2	76.8	201	7	US-10-995-561-32893	Sequence 32893, A
C 40	19.2	76.8	201	7	US-10-995-561-61135	Sequence 61135, A
C 41	19.2	76.8	201	7	US-10-995-561-61137	Sequence 61137, A
C 42	19.2	76.8	88116	7	US-10-995-561-61168	Sequence 61168, A
C 43	19.2	76.8	222094	7	US-10-995-561-13351	Sequence 13351, A
C 44	19.2	76.8	222094	7	US-10-995-561-13244	Sequence 13244, A
C 45	19.2	76.0	1125000	7	US-10-995-561-44307	Sequence 44307, A
C 46	18.6	74.4	25	8	US-10-995-561-13286	Sequence 13286, A
C 47	18.6	74.4	201	7	US-11-121-849-663017	Sequence 663017, A
C 48	18.6	74.4	201	7	US-10-995-561-35734	Sequence 35734, A
C 49	18.6	74.4	201	7	US-10-995-561-52602	Sequence 52602, A
C 50	18.6	74.4	201	7	US-10-995-561-52678	Sequence 52678, A
C 51	18.6	74.4	201	7	US-10-995-561-78552	Sequence 78552, A
C 52	18.6	74.4	201	8	US-11-124-368A-4303	Sequence 4303, Ap
C 53	18.6	74.4	16433	8	US-11-124-368A-4321	Sequence 4321, Ap
C 54	18.6	74.4	22855	8	US-11-124-368A-2875	Sequence 2875, Ap
C 55	18.6	74.4	38553	7	US-10-995-561-13477	Sequence 13477, A
C 56	18.6	74.4	35962	7	US-10-775-169-234	Sequence 234, App
C 57	18.6	74.4	38023	7	US-10-995-561-13251	Sequence 13251, A
C 58	18.6	74.4	95832	8	US-11-124-368A-23900	Sequence 23900, Ap
C 59	18.6	74.4	95832	7	US-10-995-561-13273	Sequence 13273, A
C 60	18.6	74.4	110608	8	US-11-121-086-12	Sequence 12, Appl
C 61	18.6	74.4	119160	8	US-11-121-086-24	Sequence 24, Appl
C 62	18.6	74.4	120096	8	US-11-121-086-49	Sequence 49, Appl
C 63	18.6	74.4	130472	7	US-10-995-561-13312	Sequence 13312, A
C 64	18.6	74.4	149419	8	US-11-112-908-49	Sequence 49, Appl
C 65	18.6	74.4	150038	8	US-11-121-086-23	Sequence 23, Appl
C 66	18.6	74.4	153376	8	US-11-121-086-5	Sequence 5, Appl
C 67	18.6	74.4	156260	8	US-11-121-086-87	Sequence 87, Appl
C 68	18.6	74.4	159146	8	US-11-121-086-49	Sequence 49, Appl
C 69	18.6	74.4	160170	8	US-11-121-086-32	Sequence 32, Appl
C 70	18.6	74.4	161726	8	US-11-112-908-48	Sequence 48, Appl
C 71	18.6	74.4	161726	8	US-11-112-908-52	Sequence 52, Appl
C 72	18.6	74.4	161874	8	US-11-121-086-75	Sequence 75, Appl
C 73	18.6	74.4	166111	8	US-11-112-908-47	Sequence 47, Appl
C 74	18.6	74.4	168516	8	US-11-121-086-3	Sequence 3, Appl
C 75	18.6	74.4	168656	8	US-11-112-908-59	Sequence 59, Appl
C 76	18.6	74.4	170285	8	US-11-112-908-58	Sequence 58, Appl
C 77	18.6	74.4	172147	8	US-11-112-908-22	Sequence 22, Appl
C 78	18.6	74.4	175100	8	US-11-121-086-21	Sequence 21, Appl
C 79	18.6	74.4	176503	8	US-11-121-086-53	Sequence 53, Appl
C 80	18.6	74.4	179777	8	US-11-121-086-106	Sequence 106, App
C 81	18.6	74.4	181172	8	US-11-121-086-41	Sequence 41, Appl
C 82	18.6	74.4	182303	8	US-11-121-086-45	Sequence 45, Appl
C 83	18.6	74.4	189252	8	US-11-121-086-54	Sequence 54, Appl
C 84	18.6	74.4	191684	8	US-11-121-086-2	Sequence 2, Appl
C 85	18.6	74.4	191684	8	US-11-121-086-2	Sequence 2, Appl
C 86	18.6	74.4	197096	8	US-11-121-086-107	Sequence 107, App
C 87	18.6	74.4	197781	8	US-11-112-908-34	Sequence 34, Appl
C 88	18.6	74.4	198161	7	US-10-775-169-52	Sequence 52, Appl
C 89	18.6	74.4	203467	8	US-11-121-086-50	Sequence 50, Appl
C 90	18.6	74.4	212805	8	US-11-121-086-19	Sequence 19, Appl
C 91	18.6	74.4	220895	7	US-10-775-169-88	Sequence 88, Appl
C 92	18.6	74.4	220895	7	US-10-775-169-88	Sequence 88, Appl
C 93	18.6	74.4	246960	8	US-11-121-086-8	Sequence 8, Appl
C 94	18.6	74.4	645179	7	US-10-995-561-13293	Sequence 13293, A
C 95	18.6	74.4	1080000	7	US-10-928-446A-1	Sequence 1, Appl

```
c 96 18.6 74.4 1080000 7 US-10-928-446A-1
c 97 18.6 74.4 1080000 7 US-10-928-446A-181
c 98 18.6 74.4 1080000 7 US-10-928-446A-181
c 99 18.6 74.4 1080000 7 US-10-928-446A-183
c 100 18.6 74.4 1080000 7 US-10-928-446A-183
c 101 18.6 74.4 1080000 7 US-10-928-446A-185
c 102 18.6 74.4 1080000 7 US-10-928-446A-185
c 103 18.6 74.4 1080000 7 US-10-928-446A-187
c 104 18.6 74.4 1080000 7 US-10-928-446A-187
c 105 18.6 74.4 1080000 7 US-10-928-446A-189
c 106 18.6 74.4 1080000 7 US-10-928-446A-189
c 107 18.6 74.4 1080000 7 US-10-928-446A-191
c 108 18.6 74.4 1080000 7 US-10-928-446A-191
c 109 18.6 74.4 1080000 7 US-10-928-446A-193
c 110 18.6 74.4 1080000 7 US-10-928-446A-193
c 111 18.6 74.4 1080000 7 US-10-928-446A-195
c 112 18.6 74.4 1080000 7 US-10-928-446A-195
c 113 18.6 74.4 1080000 7 US-10-928-446A-197
c 114 18.6 74.4 1080000 7 US-10-928-446A-197
c 115 18.6 74.4 1080000 7 US-10-928-446A-199
c 116 18.6 74.4 1080000 7 US-10-928-446A-199
c 117 18.6 74.4 1080000 7 US-10-928-446A-201
c 118 18.6 74.4 1080000 7 US-10-928-446A-201
c 119 18.6 74.4 1691140 8 US-11-091-018-1
c 120 18.4 73.6 218821 8 US-11-121-086-31
c 121 18.2 72.8 25 8 US-11-121-849-249082
c 122 18.2 72.8 25 8 US-11-121-849-250268
c 123 18.2 72.8 201 7 US-10-995-561-62046
c 124 18.2 72.8 201 7 US-10-995-561-62046
c 125 18.2 72.8 201 7 US-10-995-561-78402
c 126 18.2 72.8 201 7 US-11-136-527-6987
c 127 18.2 72.8 971 7 US-10-750-185-38521
c 128 18.2 72.8 971 7 US-10-750-623-38521
c 129 18.2 72.8 1182 7 US-10-750-185-38683
c 130 18.2 72.8 1182 7 US-10-750-623-38683
c 131 18.2 72.8 2063 8 US-11-136-527-2891
c 132 18.2 72.8 46752 7 US-10-995-561-13410
c 133 18.2 72.8 53331 7 US-10-995-561-13476
c 134 18.2 72.8 54767 7 US-10-995-561-13357
c 135 18.2 72.8 142605 8 US-11-121-086-64
c 136 17.8 71.2 25 8 US-11-121-849-19047
c 137 17.8 71.2 201 7 US-10-995-561-28516
c 138 17.8 71.2 201 7 US-10-995-561-44308
c 139 17.8 71.2 79666 8 US-11-043-752-1
c 140 17.8 71.2 162537 8 US-11-121-086-59
c 141 17.8 71.2 167116 8 US-11-121-086-44
c 142 17.8 71.2 222094 7 US-10-995-561-13244
c 143 17.6 70.4 25 8 US-11-121-849-27146
c 144 17.6 70.4 25 8 US-11-121-849-521577
c 145 17.6 70.4 25 8 US-11-121-849-621873
c 146 17.6 70.4 201 7 US-10-995-561-15839
c 147 17.6 70.4 201 7 US-10-995-561-27244
c 148 17.6 70.4 201 7 US-10-995-561-41593
c 149 17.6 70.4 201 8 US-11-124-368A-8230
c 150 17.6 70.4 201 8 US-11-124-368A-14085
```

Sequence 1, Appli  
Sequence 181, App  
Sequence 181, App  
Sequence 183, App  
Sequence 183, App  
Sequence 185, App  
Sequence 185, App  
Sequence 187, App  
Sequence 187, App  
Sequence 189, App  
Sequence 189, App  
Sequence 191, App  
Sequence 191, App  
Sequence 193, App  
Sequence 193, App  
Sequence 195, App  
Sequence 195, App  
Sequence 197, App  
Sequence 197, App  
Sequence 199, App  
Sequence 199, App  
Sequence 201, App  
Sequence 201, App  
Sequence 1, Appli  
Sequence 31, Appl  
Sequence 249082,  
Sequence 250268,  
Sequence 62046, A  
Sequence 78402, A  
Sequence 6987, Ap  
Sequence 38521, A  
Sequence 38521, A  
Sequence 38683, A  
Sequence 38683, A  
Sequence 2891, Ap  
Sequence 13410, A  
Sequence 13476, A  
Sequence 13357, A  
Sequence 64, Appl  
Sequence 23, Appl  
Sequence 19047, A  
Sequence 28516, A  
Sequence 44308, A  
Sequence 1, Appli  
Sequence 59, Appl  
Sequence 44, Appl  
Sequence 13244, A  
Sequence 27146, A  
Sequence 521577,  
Sequence 621873,  
Sequence 15839, A  
Sequence 27244, A  
Sequence 41593, A  
Sequence 8230, Ap  
Sequence 14085, A

us-10-995-561-79948/c  
; SEQ ID NO 79948  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-79948

Query Match 93.6%; Score 23.4; DB 7; Length 201;  
Best Local Similarity 96.0%; Pred. No. 0.14; 1; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0

QY 1 AAAAAAAAAATCAGCCGCGCATGG 25  
|||||  
Db 65 AAAAAAAAAATCAGCCGCGCATGG 41  
|||||

RESULT 2  
US-10-995-561-13487/c  
; Sequence 13487, Application US/109955561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13487  
; LENGTH: 41309  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13487

Query Match 93.6%; Score 23.4; DB 7; Length 41309;  
Best Local Similarity 96.0%; Pred. No. 0.44; 1; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0

QY 1 AAAAAAAAAATCAGCCGCGCATGG 25  
|||||  
Db 19143 AAAAAAAAAATCAGCCGCGCATGG 19119  
|||||

RESULT 3  
US-10-995-561-72586/c  
; Sequence 72586, Application US/109955561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72586  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-72586

Query Match 81.6%; Score 20.4; DB 7; Length 201;  
Best Local Similarity 95.5%; Pred. No. 3.3;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAATCAGCCGCGGCA 22  
|||||  
Db 109 AAAAAAAAAATCAGCCGCGGCA 88  
|||||

## ALIGNMENTS

RESULT 1  
US-10-995-561-79948/c  
; Sequence 79948, Application US/109955561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702



```

RESULT 4
US-10-995-561-84366/c
; Sequence 84366, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 84366
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-84366

Query Match      81.6%; Score 20.4; DB 7; Length 201;
Best Local Similarity 95.5%; Pred. No. 3.3;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCAGGCCAGGCA 22
      |||||||
DB      109 AAAAAAAAAATCAGGCCAGGCA 88

RESULT 5
US-10-995-561-13294/c
; Sequence 13294, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13294
; LENGTH: 15804
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(15804)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13294

Query Match      81.6%; Score 20.4; DB 7; Length 15804;
Best Local Similarity 95.5%; Pred. No. 8.3;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCAGGCCAGGCA 22
      |||||||
DB      58 AAAAAAAAAATCAAGGCCAGGCA 37

RESULT 6
US-10-995-561-13436/c
; Sequence 13436, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559

```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-27

Query Match      81.6%; Score 20.4; DB 8; Length 153142;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCA 22
   |||||
Db 63127 AAAAAAAAAATCAGGCCAGGCA 63106

RESULT 9
US-10-995-561-33166/c
; Sequence 33166, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33166
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-33166

Query Match      80.8%; Score 20.2; DB 7; Length 201;
Best Local Similarity 88.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 76 AAAAAAAAAATCAGGCCAGGCATGG 52

RESULT 12
US-10-995-561-13352
; Sequence 13352, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13352
; LENGTH: 32157
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13352

Query Match      80.8%; Score 20.2; DB 7; Length 32157;
Best Local Similarity 88.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 12697 AAAAAAAAAATGTGGCCAGGCATGG 12721

RESULT 13
US-11-124-368A-2910
; Sequence 2910, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-27

Query Match      81.6%; Score 20.4; DB 8; Length 153142;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCA 22
   |||||
Db 63127 AAAAAAAAAATCAGGCCAGGCA 63106

RESULT 9
US-10-995-561-33166/c
; Sequence 33166, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33166
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-33166

Query Match      80.8%; Score 20.2; DB 7; Length 201;
Best Local Similarity 88.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 201 ACAAAAAAATTAAGGCCAGGCATGG 177

RESULT 10
US-10-995-561-50016/c
; Sequence 50016, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50016
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-50016

Query Match      80.8%; Score 20.2; DB 7; Length 201;
Best Local Similarity 88.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGGCCAGGCATGG 25
   |||||
Db 176 AAAAAAAAAAATTAAGGCCAGGCATGG 152

RESULT 11
US-10-995-561-56961/c
```

```

; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2910
; LENGTH: 44229
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2910

```

```
Query Match      80.8%; Score 20.2; DB 8; Length 44229;
Best Local Similarity 88.0%; Pred. No. 13;
Matches 22: Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

**Qy**            1 AAAAAAAAAATCAGGCCAGGCATGG 25  
             ||||| | | | | | | | | |  
**Dh**            41383 AAAATAACTCAGGCCAGGCATGG 41407

RESULT 14  
US-10-995-561-13359  
; Sequence 13359, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 13359
; LENGTH: 60844
; TYPE: DNA
; ORGANISM: Homo sapiens
ITS-10-995-561-13359

```

Query Match	80.8%	Score 20.2;	DB 7;	Length 60844;
Best Local Similarity	88.0%;	Pred. No. 14;		
Matches 22:	Conservative	0;	Mismatches 3;	Indels 0;
	Gaps	0;		

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
||||| ||||| ||||| |||||  
nb 53002 AAAACAAGCAAGGCCAGGCATGG 53026

```

RESULT 15
US-11-124-368A-2911
; Sequence 2911, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2911
; LENGTH: 66280
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2911

```

Query Match	80.8%;	Score 20.2;	DB 8;	Length 66280;
Best Local Similarity	88.0%;	Pred. No. 14;		

	Matches	22; Conservative	0; Mismatch
Qy	1	AAAAAAAAATCAGGCCAGGCATGG	25
Dp	10801	AAAAATAACTCAGGCCAGGCATGG	10825

RESULT 16  
US-10-857-780-5  
; Sequence 5, Application US/10857780  
; Publication No. US20050272043A1  
; GENERAL INFORMATION:  
; APPLICANT: ROTH, RICHARD B.  
; APPLICANT: BRAUN, ANDREAS  
; APPLICANT: KAMMERER, STEFAN M.  
; APPLICANT: NELSON, MATTHEW ROBERTS  
; APPLICANT: RENE LAND, RIKARD HENRY  
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: SEQ-4069-CP  
; CURRENT APPLICATION NUMBER: US/10/857, 780

```

: CURRENT FILING DATE: 2004-05-05
: PRIORITY APPLICATION NUMBER: 10/7
: PRIOR FILING DATE: 2003-11-25
: PRIORITY APPLICATION NUMBER: 60/4
: PRIOR FILING DATE: 2003-07-24
: PRIORITY APPLICATION NUMBER: 60/5
: PRIOR FILING DATE: 2003-11-25
: NUMBER OF SEQ ID NOS: 4962
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 5

```

```

, LENGTH: 86950
, TYPE: DNA
, ORGANISM: Homo sapiens
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (20282)..(20282)
, OTHER INFORMATION: n is a, c, g, or t
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (39985)..(39985)
, OTHER INFORMATION: n is a, c, g, or t
, FEATURE:
, NAME/KEY: misc feature
, LOCATION: (39989)..(39989)
, OTHER INFORMATION: n is a, c, g, or t
, FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: (51996)..(51996)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
;
; NAME/KEY: misc feature
; LOCATION: (64732)..(64732)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
;
; NAME/KEY: misc feature
; LOCATION: (83916)..(83919)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
;
; NAME/KEY: misc feature
; LOCATION: (86196)..(86196)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-857-780-5

```

Query Match	80.8%	Score 20.2;	DB 7;	Length 86950;
Best Local Similarity	88.0%;	Pred. No. 15;		
Matches 22: Conservative	0;	Mismatches	3;	Indels 0;
				Gaps 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
|||  
26574 AAAAAAAAAAAAAAAAAAGGCCAGGCATGG 26598

```

; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13262
; LENGTH: 141121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1...141121)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables :
; US-10-995-561-13262

```

```
Query Match      80.8%; Score 20.2; DB 7; Length 141121;
Best Local Similarity 88.0%; Pred. No. 16;
Matches 22: Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY	1	AAAAAAAAATCAGGCCAGGCATGG	25
pB	82982	ACAAAAAATTAAAGGCCAGGCATGG	82

```

RESULT 20
US-11-121-086-29/c
; Sequence 29, Application US/11121086
; Publication NO. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent version 3.3
; SEQ ID NO 29
; LENGTH: 160226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-29

```

Query Match 80.8%; Score 20.2; DB 8; Length 160226;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0;

Qy 1 AAAAAAAAAATCAGGCCAGGCATGG 25  
||||| | | | | | | | | |  
Dh 147899 AAAAAAAAAAAAAAGGCCAGGCATGG 14

```

RESULT 21
US-11-112-908-59
Sequence 59, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE OF INVENTION: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIORITY APPLICATION NUMBER: US 60/564,758
PRIORITY FILING DATE: 2004-04-23
PRIORITY APPLICATION NUMBER: US 60/575,978
PRIORITY FILING DATE: 2004-06-01
PRIORITY APPLICATION NUMBER: US 60/631,702
PRIORITY FILING DATE: 2004-11-30
PRIORITY APPLICATION NUMBER: US 60/633,826

```

;; PRIOR FILING DATE: 2004-12-07  
;; NUMBER OF SEQ ID NOS: 511  
;; SOFTWARE: PatentIn version 3.3  
;; SEQ ID NO 59  
;; LENGTH: 168656  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-11-112-908-59

Query Match 80.8%; Score 20.2; DB 8; Length 168656;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 159496 AAAAAAAAAATCCTGCGCCAGGCATGG 159520

RESULT 22  
US-11-112-908-58  
; Sequence 58, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 58  
; LENGTH: 170285  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-58

Query Match 80.8%; Score 20.2; DB 8; Length 170285;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 159501 AAAAAAAAAATCCTGCGCCAGGCATGG 159525

RESULT 23  
US-11-121-086-67/c  
; Sequence 67, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 67  
; LENGTH: 179666  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-11-121-086-67

Query Match 80.8%; Score 20.2; DB 8; Length 179666;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 168855 AAAAAAAAAAAGCGCCAGGCATGG 168831

RESULT 24  
US-11-112-908-39/c  
; Sequence 39, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 39  
; LENGTH: 179892  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-39

Query Match 80.8%; Score 20.2; DB 8; Length 179892;  
Best Local Similarity 88.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCAGCGCCAGGCATGG 25  
Db 101870 AAAAAAAAAACAGCGCGCATGG 101846

RESULT 25  
US-10-995-561-13274  
; Sequence 13274, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13274  
; LENGTH: 415117  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(415117)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-995-561-13274

Query Match 80.8%; Score 20.2; DB 7; Length 415117;  
Best Local Similarity 88.0%; Pred. No. 20;

```
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 268181 AAAAAAAAAAAAAAGCCAGGCATGG 268205

RESULT 26
US-10-995-561-13274/c
; Sequence 13274, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13274
; LENGTH: 415117
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(415117)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13274

Query Match 80.8%; Score 20.2; DB 7; Length 415117;
Best Local Similarity 88.0%; Pred. No. 20;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 289065 AAAAAAAAAAAAAAGCCAGGCATGG 289041

RESULT 27
US-10-995-561-36951/c
; Sequence 36951, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36951
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-36951

Query Match 79.2%; Score 19.8; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 6.2;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 104 AAAAAAAAAAAAAAGCCAGGCATGG 80

RESULT 28
US-10-995-561-50172/c
; Sequence 50172, Application US/10995561
; Publication No. US20050272054A1
```

```
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50172
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-50172

Query Match 79.2%; Score 19.8; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 6.2;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCATGG 25
Db 122 AAAAAAAAAAAAAATGCCAGGCRTGG 98

RESULT 29
US-11-121-086-35
; Sequence 35, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 170995
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-35

Query Match 79.2%; Score 19.8; DB 8; Length 170995;
Best Local Similarity 91.3%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCAGCGCCAGGCAT 23
Db 111971 AAAAAAAAAATCATGCCAGGCAT 111993

RESULT 30
US-10-995-561-13303/c
; Sequence 13303, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13303
; LENGTH: 201990
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

RESULT 33
US-11-121-849-250266
; Sequence 250266, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121.849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 250266
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-250266

Query Match          76.8%; Score 19.2; DB 8; Length 25;
Best Local Similarity 87.5%; Pred. No. 7.4;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||| ||||| ||||| ||||| |||
Db 2 AAAACAAATCATGGCCAGGCCTG 25
   ||| ||||| ||||| ||||| |||

RESULT 34
US-11-121-849-250267
; Sequence 250267, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121.849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 250267
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-250267

Query Match          76.8%; Score 19.2; DB 8; Length 25;
Best Local Similarity 87.5%; Pred. No. 7.4;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||| ||||| ||||| ||||| |||
Db 1 AAAACAAATCATGGCCAGGCCTG 24
   ||| ||||| ||||| ||||| |||

RESULT 35
US-10-995-561-28311
; Sequence 28311, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561

```



; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 28311  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-28311  
  
 Query Match 76.8%; Score 19.2; DB 7; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 AAAAAAAAAATCAGGCCAGGCATG 24  
 |||||  
 Db 134 AAAAAAAAAAGGCCAGGCATG 157  
  
 RESULT 36  
 US-10-995-561-28314  
 ; Sequence 28314, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 28314  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-28314  
  
 Query Match 76.8%; Score 19.2; DB 7; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 AAAAAAAAAATCAGGCCAGGCATG 24  
 |||||  
 Db 60 AAAAAAAAAAGGCCAGGCATG 83  
  
 RESULT 37  
 US-10-995-561-28356  
 ; Sequence 28356, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 28356  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-28356  
  
 Query Match 76.8%; Score 19.2; DB 7; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 AAAAAAAAAATCAGGCCAGGCATG 24  
 |||||  
 Db 60 AAAAAAAAAAGGCCAGGCATG 83  
  
 RESULT 38  
 US-10-995-561-32893/c  
 ; Sequence 32893, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 32893  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-32893  
  
 Query Match 76.8%; Score 19.2; DB 7; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 AAAAAAAAAATCAGGCCAGGCATG 24  
 |||||  
 Db 24 AAAAAAAAAATTAAGGCCAGGCATG 1  
  
 RESULT 39  
 US-10-995-561-61135/c  
 ; Sequence 61135, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 61135  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-61135  
  
 Query Match 76.8%; Score 19.2; DB 7; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 AAAAAAAAAATCAGGCCAGGCATG 24  
 |||||  
 Db 68 AAAAAAAAAAGGCCAGGCATG 45  
  
 RESULT 40  
 US-10-995-561-61137/c  
 ; Sequence 61137, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 61137  
 ; LENGTH: 201  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-995-561-61137

```
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 61137
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-61137

Query Match          76.8%; Score 19.2; DB 7; Length 201;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||||| ||||| ||||| ||||| |||||
Db 142 AAAAAAAAAAAGGCCAGGCATG 119

RESULT 41
US-10-995-561-61168/c
/ Sequence 61168, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 61168
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-61168

Query Match          76.8%; Score 19.2; DB 7; Length 201;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||||| ||||| ||||| ||||| |||||
Db 113 AAAAAAAAAAAGGCCAGGCATG 90

RESULT 42
US-10-995-561-13351/c
/ Sequence 13351, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13351
/ LENGTH: 88116
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-13351

Query Match          76.8%; Score 19.2; DB 7; Length 88116;
Best Local Similarity 87.5%; Pred. No. 42;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||||| ||||| ||||| ||||| |||||
Db 55029 AAAAAAAAAAAGGCCAGGCATG 55006

RESULT 43
US-10-995-561-13244
/ Sequence 13244, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13244
/ LENGTH: 222094
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-13244

Query Match          76.8%; Score 19.2; DB 7; Length 222094;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCAGCGCCAGGCATG 24
   ||||| ||||| ||||| ||||| |||||
Db 163769 AAAAAAAAAAAGGCCAGGCATG 163792

RESULT 44
US-10-995-561-44307/c
/ Sequence 44307, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 44307
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-44307

Query Match          76.0%; Score 19; DB 7; Length 201;
Best Local Similarity 90.3%; Pred. No. 14;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAAAAATCAGCGCCAGGCATGG 25
   ||||| ||||| ||||| ||||| |||||
Db 118 AAAAAATCAGCGCCAGGCATGG 98

RESULT 45
US-10-995-561-13286/c
/ Sequence 13286, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13286
; LENGTH: 1125000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1125000)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13286

Query Match          76.0%; Score 19; DB 7; Length 1125000;
Best Local Similarity 90.5%; Pred. No. 83;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      5 AAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
Db      931121 AAAAAATCAGCGCCAGGCRTGG 931101

RESULT 46
US-11-121-849-663017
; Sequence 663017, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; TITLE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 663017
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-663017

Query Match          74.4%; Score 18.6; DB 8; Length 25;
Best Local Similarity 84.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAATCAGCGCCAGGCATGG 25
      ||| ||||| ||||| ||||| |||||
Db      1 AAACAATATGAGGGCCAGGCATGG 25

RESULT 47
US-10-995-561-35734/c
; Sequence 35734, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35734
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-35734

Query Match          74.4%; Score 18.6; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
Db      150 AAAAAAAATCAGCGCCAGGCATGG 126

RESULT 48
US-10-995-561-52602/c
; Sequence 52602, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52602
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-52602

Query Match          74.4%; Score 18.6; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
Db      130 AAAAAAAATCAGCGCCAGGCATGG 106

RESULT 49
US-10-995-561-52678/c
; Sequence 52678, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52678
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-52678

Query Match          74.4%; Score 18.6; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAATCAGCGCCAGGCATGG 25
      ||||| ||||| ||||| ||||| |||||
Db      84 AAAAAAAATCAGCGCCAGGCATGG 60

RESULT 50
US-10-995-561-78552/c
; Sequence 78552, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
```

Mon Feb 6 12:23:11 2006

```
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 78552
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-78552

Query Match      74.4%; Score 18.6; DB 7; Length 201;
Best Local Similarity 84.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCAGGCCAGGCATGG 25
        |||||
Db      73 AAAAAAAAAAATAGCCAGGCATGG 49
```

Search completed: February 3, 2006, 16:19:44  
Job time : 362.111 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-3

Perfect score: 25

Sequence: 1 aaaaacaaccaccactgagctggg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_ov.\*

5: gb\_om.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_sts.\*

11: gb\_ey.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	4192	9	BC080307 Mus muscu
2	25	100.0	4618	9	BC059216 Mus muscu
3	25	100.0	211414	9	AC104885 Mus muscu
4	21.8	87.2	100983	9	AL663075 Mouse DNA
5	21.8	87.2	169405	9	AL596256 Mouse DNA
6	21.8	87.2	182785	8	AC145937 Pan trogl
7	21.8	87.2	214993	14	AC055773 Mus muscu
8	21.8	87.2	276234	14	AC112380 Rattus no
9	20.8	83.2	182955	4	AC149756 Bos tauru
10	20.2	80.8	2526	5	BC063914 Xenopus t
11	20.2	80.8	2588	8	BC075515 Human DNA
12	20.2	80.8	114676	8	BX323048 Human DNA
13	20.2	80.8	128000	8	AC098690 Homo sapi
14	20.2	80.8	128638	8	AC015968 Homo sapi
15	20.2	80.8	145569	5	AL935194 Zebrafish
16	20.2	80.8	155263	14	AC053517 Homo sapi
17	20.2	80.8	159976	14	AC016124 Homo sapi
18	20.2	80.8	173256	14	AC073301 Homo sapi

AC013557 Homo sapi	14	AC013557	80.8	178806	14	AC013557
AC104828 Homo sapi	8	AC104828	80.8	181489	8	AC104828
AC099068 Rattus no	14	AC099068	80.8	243279	14	AC099068
AC095147 Rattus no	14	AC095147	80.8	256914	14	AC095147
AC160578 Bos tauru	14	AC160578	80.8	290052	14	AC160578
CR854901 Danio rer	14	CR854901	79.2	192168	14	CR854901
BX469934 Zebrafish	5	BX469934	79.2	225586	5	BX469934
AC098538 Rattus no	14	AC098538	79.2	255707	14	AC098538
M84354 Gallus gall	5	CHRYPEX	77.6	1288	5	CHRYPEX
X60778 G.gallus DN	5	GWALTER	77.6	1288	5	GWALTER
BX548018 Zebrafish	5	BX548018	77.6	113040	5	BX548018
AC153985 Mus muscu	14	AC153985	77.6	169194	14	AC153985
AL953906 Mouse DNA	9	AL953906	77.6	172246	9	AL953906
AC153907 Mus muscu	9	AC153907	77.6	196286	9	AC153907
CR847836 Danio rer	14	CR847836	77.6	205309	14	CR847836
CR545469 Danio rer	14	CR545469	77.6	214525	14	CR545469
AC105669 Rattus no	14	AC105669	77.6	261112	14	AC105669
AF302834 Haliotis	2	AF302834	76.8	836	2	AF302834
AL627225 Human DNA	8	AL627225	76.8	3994	8	AL627225
AL104180 Leishmani	14	AL104180	76.8	90100	14	AL104180
AL359978 Homo sapi	14	AL359978	76.8	110000	14	AL359978
Continuation (2 of	14	AL359978_1	76.8	110000	14	AL359978_1
Continuation (3 of	14	BX936368_2	76.8	110000	14	BX936368_2
CT005261 Leishmani	14	CT005261_1	76.8	110000	14	CT005261_1
Continuation (2 of	8	AL353806	76.8	111051	8	AL353806
AL353806 Human DNA	14	AL353806	76.8	124350	14	AL353806
AC129011 Leishmani	8	AC129011	76.8	134793	8	AC129011
AC020922 Homo sapi	14	AC020922	76.8	138861	14	AC020922
AC154498 Mus muscu	14	AC154498	76.8	141663	14	AC154498
AC145018 Felis cat	14	AC145018	76.8	141663	14	AC145018
AC157291 Bos tauru	14	AC157291	76.8	154763	14	AC157291
AC016868 Homo sapi	8	AC016868	76.8	155691	8	AC016868
AC154788 Mus muscu	14	AC154788	76.8	166692	14	AC154788
CR354558 Zebrafish	5	CR354558	76.8	177206	5	CR354558
AC147399 Canis fam	14	AC147399	76.8	178483	14	AC147399
AL591807 Human Chr	8	CNS07EQ	76.8	183418	8	CNS07EQ
AC122209 Mus muscu	9	AC122209	76.8	185883	9	AC122209
AL611949 Mouse DNA	5	AL611949	76.8	186276	5	AL611949
BX548158 Zebrafish	5	BX548158	76.8	190936	5	BX548158
AL603708 Mouse DNA	9	AL603708	76.8	198743	9	AL603708
AC073687 Mus muscu	14	AC073687	76.8	201331	14	AC073687
AL845296 Mouse DNA	9	AL845296	76.8	203414	9	AL845296
CT009761 Mus muscu	14	CT009761	76.8	208369	14	CT009761
AC162504 Bos tauru	14	AC162504	76.8	217178	14	AC162504
AC123563 Rattus no	14	AC123563	76.8	219934	14	AC123563
AC144469 Canis fam	14	AC144469	76.8	223041	14	AC144469
AL732573 Mus muscu	14	AL732573	76.8	223496	14	AL732573
AC155156 Bos tauru	14	AC155156	76.8	224690	14	AC155156
AC154520 Mus muscu	14	AC154520	76.8	231513	14	AC154520
AC096842 Rattus no	14	AC096842	76.8	232927	14	AC096842
AC140246 Mus muscu	14	AC140246	76.8	250770	14	AC140246
AC105710 Rattus no	14	AC105710	76.8	280115	14	AC105710
AC135235 Mus muscu	14	AC135235	76.8	335638	14	AC135235
AC131401 Rattus no	14	AC131401	76.8	338015	14	AC131401
BX537351 Zebrafish	5	BX537351	76.0	161005	5	BX537351
BV421538 S229P6239	8	BV421538	75.2	510	8	BV421538
M37920 Homo sapien	10	HUMAMP001	75.2	673	10	HUMAMP001
BV474964 G591P6285	10	BV474964	75.2	698	10	BV474964
BV474965 G591P6285	10	BV474965	75.2	735	10	BV474965
BV485979 S215P6074	10	BV485979	75.2	736	10	BV485979
BV474961 G591P6285	10	BV474961	75.2	775	10	BV474961
BV493818 S217P6626	10	BV493818	75.2	775	10	BV493818
AC100102 Mus muscu	14	AC100102	75.2	44437	14	AC100102
AL441923 Human DNA	8	AL441923	75.2	76656	8	AL441923
AC008590 Homo sapi	14	AC008590	75.2	98692	14	AC008590
AC113637 Rattus no	8	AC113637_0	75.2	110000	8	AC113637_0
Continuation (3 of	14	AC113637_2	75.2	110000	14	AC113637_2
AC118389 Rattus no	14	AC118389_0	75.2	110000	14	AC118389_0
Continuation (2 of	14	AC118389_1	75.2	110000	14	AC118389_1
Continuation (13 of	15	AP008212_129	75.2	110000	15	AP008212_129
AP003619 Oryza sat	15	AP003619	75.2	126757	15	AP003619
AC149475 Zea mays	14	AC149475	75.2	126765	14	AC149475
AC151050 Zea mays	14	AC151050	75.2	134692	14	AC151050





## CDS

276. .1310  
 /gene="Tmem44"  
 /codon\_start=1  
 /product="Tmem44 protein"  
 /protein\_id="AAH80307.1"  
 /db\_xref="GI:51330402"  
 /db\_xref="GeneID:224090"  
 /db\_xref="MGI:1924489"  
 /translations="MNPMTLPVCGSKSKSGSSRRRRHRLRSVPALALPLSL  
 GPWATWAIKPAIPVRGQORLLGSLQENPEVGYLLGALAGSWSARIPPSN  
 ICKGSLSYHLWTRLSALGLLYASAIVAHQRQPEYLLQATPWLFI SLGRALDLA  
 IIFLSCVKSIMRGRFATFAEARSADQALITCAKEEENQEARTEKNSDWLPT  
 SLSHCKPLRTMTAISRYMELTIRPAQQAGCATRLPGDGTSTRDAASQPSYPIQ  
 VIOARVSSSSSSSVSSINSLSLEQKYWEALNSEQWDPEDVNLKRKDGELLYVQVRRAS  
 LSPVDLASDD"

## ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 4192;  
 Best Local Similarity 100.0%; Pred. No. 0.33;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGAGCTGGG 25

Db 3769 AAAAAACACCACTGAGCTGGG 3793

## RESULT 2

BC059216  
 LOCUS Mus musculus transmembrane protein 44, mRNA linear ROD 01-JUN-2005  
 IMAGE:6408767)

ACCESSION BC059216.1 GI:37747503

## KEYWORDS

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,  
 Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S.,  
 Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,  
 Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J.,  
 McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,  
 Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., R.A.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Heltón E., Kettman M., Madan A., Rodríguez S.,  
 Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,  
 Bouffard M.C., Blakeley R.W., Touchman J.W., Green E.D.,  
 Dickson M.C., Rodríguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butlerfield Y.S., Krzywiński M.I., Skalska U., Smalios D.E.,  
 Schnerch A., Schein J.E., Jones S.J. and Marra M.A.

## CONSRM

TITLE Mammalian Gene Collection Program Team  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences

## JOURNAL

PUBMED 12477932

2 (bases 1 to 4618)

## REFERENCE

AUTHORS NIH MGC Project

TITLE Direct Submission

## JOURNAL

Submitted (01-OCT-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Bethesda, MD 20892-2590, USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [ccapbs-remail.nih.gov](mailto:ccapbs-remail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

CDNA Library Preparation: M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 125 Row: 9 Column: 11.

## FEATURES

## source

1. .4618  
 /Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6408767"  
 /tissue="Brain, enriched mouse brain 12.5dp"  
 /clone\_lib="NIH BMAP\_F00"  
 /lab\_host="DH10B"  
 /note="Vector: pYX-ASC"

## ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 4618;  
 Best Local Similarity 100.0%; Pred. No. 0.33;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGAGCTGGG 25

Db 4180 AAAAAACACCACTGAGCTGGG 4204

## RESULT 3

AC104885/c  
 LOCUS Mus musculus chromosome 16, clone RP23-125P19, complete sequence.  
 DEFINITION AC104885  
 ACCESSION AC104885.8 GI:38198721

## VERSION

HTG.

## KEYWORDS

Mus musculus (house mouse)

## SOURCE

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

## TITLE

1 (bases 1 to 211414)

## JOURNAL

REFERENCE

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., DeAstellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
 Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., MCPheeters, R., Meldrum, J.,  
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.



```

repeat_region complement(12158..12455)
repeat_region /rpt_family="RMER19B"
repeat_region complement(12460..13065)
repeat_region /rpt_family="RMER19B"
repeat_region 13066..13254
repeat_region /rpt_family="B2 Mmi"
repeat_region complement(13255..13285)
repeat_region /rpt_family="RMER19B"
repeat_region complement(13520..13652)
repeat_region /rpt_family="RMER15"
repeat_region complement(13736..13895)
repeat_region /rpt_family="RLTR11A"
repeat_region complement(14020..14224)
repeat_region /rpt_family="BGLII"
repeat_region complement(14225..14325)
repeat_region /rpt_family="RMER15"
repeat_region 14712..14733

Query Match 100.0%; Score 25; DB 9; Length 211414;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGG 25
Db 58978 AAAAAACACCACTGAGCTGG 58954
|||||
1 AAAAAACACCACTGAGCTGG 25

RESULT 4
AL663075/c
LOCUS Mouse DNA sequence from clone RP23-430121 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL663075
VERSION AL663075.6 GI:19572106
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 100983)
Direct Submission
Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi:19031807.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
RP23-430121 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.

```

```

FEATURES
source
Location/Qualifiers
1..100983
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="11"
/clone_lib="RPCI-23"
2761..72904
/locus_tag="RP23-430121.3-001"
pseudo
2761..72904
/locus_tag="RP23-430121.3-001"
/notes="match: proteins: Sw:Q28894 Sw:Q81UE2"
pseudo
/codon_start=1
/product="extracellular proteinase inhibitor (Expi)
pseudogene"
5461..7738
/genes="RP23-430121.1"
/locus_tag="RP23-430121.1-001"
join(5461..5643,6264..6412,7538..7738)
/genes="RP23-430121.1"
/locus_tag="RP23-430121.1-001"
/product="novel protein similar to extracellular
proteinase inhibitor Expi"
/notes="match: ESTs: Em:AA201047.1 Em:AA237876.1
Em:AA244943.1 Em:AA274581.1 Em:AA561990.1 Em:AA666504.1
Em:AA981470.1 Em:BE624328.1 Em:CB319028.1 Em:W08116.1
match: cDNAs: Em:BC010986.1 Em:BC038870.1"
join(5562..5643,6264..6412)
/genes="RP23-430121.1"
/locus_tag="RP23-430121.1-001"
/standard_name="OTTMUSP0000000924"
/notes="match: proteins: Sw:PI4730 Sw:Q62477 Tr:Q91VQ6
Tr:Q99M36"
/codon_start=1
/product="novel protein similar to extracellular
proteinase inhibitor Expi"
/protein_id="CAI25670.1"
/db_xref="GI:56206511"
/db_xref="InterPro:IPRO008197"
/translation="MKTATVFLVALITVGMNTTYVYVSCPKEFKGACPKPSPESVG
ICVDQSGGSGCPGNMKCCNSCGHCKTPVF"
7119..7724
/genes="RP23-430121.1"
/locus_tag="RP23-430121.1-001"
7738
/genes="RP23-430121.1"
/locus_tag="RP23-430121.1-001"
join(10485..10669,11327..11469,12626..12810)
/locus_tag="RP23-430121.2-001"
join(10485..10669,11327..11469,12626..12810)
/locus_tag="RP23-430121.2-001"
/product="extracellular proteinase inhibitor"
/notes="match: ESTs: Em:AW08862.1 Em:BB823505.1
Em:BB828561.1 Em:BB835917.1 Em:BE282859.1 Em:BE553674.1
Em:BE554305.1 Em:BF181080.1 Em:BG084226.2 Em:BG863639.1
Em:BI111742.1 Em:BI453987.1 Em:BI660358.1 Em:CB598051.1
match: cDNAs: Em:AF159701.1 Em:BC002038.1 Em:BC010986.1
Em:X93037.1"
join(10588..10669,11327..11469)
/locus_tag="RP23-430121.2-001"
/standard_name="OTTMUSP0000000922"
/notes="match: proteins: Sw:Q62477 Tr:Q91VQ6 Tr:Q99M36"
/codon_start=1
/product="extracellular proteinase inhibitor"
/protein_id="CAI25671.1"
/db_xref="GI:56206512"
/db_xref="InterPro:IPRO008197"
/translation="MKTATVFLVALITVGMNTTYVYVSCPKEFKGACPKPSPESVG
DERCTGDGSGGNMKCCNSCGHCKTPVF"

```

polyA\_signal

polyA\_site

gene

mRNA

CDS



```
gene
/clone_lib="RPCI-23"
join(complement(44425..44567),complement(42239..42473),
complement(32825..32993),complement(18848..18993),
complement(17953..18150),complement(16133..16327),
complement(15412..15574),complement(14767..14877),
complement(6313..6996),AL592065.7:6124..6337,
AL592065.7:7974..8052,AL592065.7:11381..11502,
AL592065.7:13481..13571,AL592065.7:13692..13906,
AL592065.7:14190..14386,AL592065.7:15840..16756,
AL592065.7:23832..23993,AL592065.7:25880..26103,
AL592065.7:28146..28602,
AL592065.7:26756..26945,AL592065.7:28146..28602,
AL592065.7:29704..29923,AL592065.7:31094..31285,
AL592065.7:31613..31836,AL592065.7:35301..35443,
AL592065.7:36114..36272,AL592065.7:36381..36566,
AL592065.7:38068..38216,AL592065.7:39489..39662,
AL592065.7:42473..42573,AL592065.7:43918..48003)
/locus_tag="RP23-5084.3-001"
join(complement(44425..44567),complement(42239..42473),
complement(32825..32993),complement(18848..18993),
complement(17953..18150),complement(16133..16327),
complement(15412..15574),complement(14767..14877),
complement(6313..6996),AL592065.7:6124..6337,
AL592065.7:7974..8052,AL592065.7:11381..11502,
AL592065.7:14190..14386,AL592065.7:15840..16756,
AL592065.7:23832..23993,AL592065.7:25880..26103,
AL592065.7:28146..28602,
AL592065.7:26756..26945,AL592065.7:28146..28602,
AL592065.7:29704..29923,AL592065.7:31094..31285,
AL592065.7:31613..31836,AL592065.7:35301..35443,
AL592065.7:36114..36272,AL592065.7:36381..36566,
AL592065.7:38068..38216,AL592065.7:39489..39662,
AL592065.7:42473..42573,AL592065.7:43918..48003)
/locus_tag="RP23-5084.3-001"
/product="novel protein"
/note="match: ESTs: Em:BB152232.2 Em:BB329744.1
Em:BB630436.1 Em:BG091148.1 Em:BU924606.1 Em:CA532211.1
Em:CA890306.1 Em:CD349776.1 Em:CD551607.1 Em:CF163861.1
Em:CF741768.1 Em:CF744131.1 Em:CF892872.1 Em:CF906293.1
match: CDNA8: Em:AB011165.2 Em:AF117754.1"
join(complement(44425..44900),complement(42239..42473),
complement(32825..32993),complement(18848..18993),
complement(17953..18150),complement(16133..16327),
complement(15412..15574),complement(14767..14877),
complement(6313..6996),AL592065.7:6124..6337,
AL592065.7:7974..8052,AL592065.7:11381..11502,
AL592065.7:13481..13571,AL592065.7:13692..13906,
AL592065.7:14190..14386,AL592065.7:15840..16756,
AL592065.7:23832..23993,AL592065.7:25880..26103,
AL592065.7:28146..28602,
AL592065.7:26756..26945,AL592065.7:28146..28602,
AL592065.7:29704..29923,AL592065.7:31094..31285,
AL592065.7:31613..31836,AL592065.7:35301..35443,
AL592065.7:36114..36272,AL592065.7:36381..36566,
AL592065.7:38068..38216,AL592065.7:39489..39662,
AL592065.7:42473..42573,AL592065.7:43918..44050)
/locus_tag="RP23-5084.3-001"
/standard_name="OTTMUSP0000001175"
/note="match: proteins: Sw:Q9UHV7 Tr:BA225519 Tr:BAC97978
Tr:Q8UTM3 Tr:Q810K4 Tr:Q8C2D6 Tr:Q9H8C0 Tr:Q9NSY9
Tr:Q9UFD8 Tr:Q9UPX5"
/codon_start=1
/product="novel protein"
/db_xref="GI:56205831"
/translation="MSSFVNSGASLEDCNLCFLADLTGKIKRWVWQPTSPAIL
FPVTEEDILSFRCCLKADVLGVRDRDQRRRELWIFWGDKNFADLTHDLSEE
EDWVNGLSYECLTLFKAHNLRLCLNRNFRWIKNFVKEVDEKPIKNSHEL
SCSTFELHGNVCTSVENIQHPVYLLSEHVTLAQONSFPVITLSPPGLNGTLT
QQAQMSDATKLLIGWKQFYPI SCGLKENSEKQDDMDWEDSLAIVEVLVAGVEM
IYPCFVLVPSQDILPAPSSVASHCSACSLGIHQVPASTRDPAMSVTLPPSPREV
QTVPDQSAQKWKVFSVSDGSTDSTSHGGKILPRKLANHVVDVWQECNNRQNRK
KYSATSGLCEETADKIGDWFVETQRTSCSLRHKSLKTRTGQAGQSGQQOQ
QVLPKHTNKKQDKSEKPKRPLTPFHRRVSVSDIEMGDTDSASQRLVISAADSQVFP
SNIRTDVAKTPQMHGTELANSQPPLPSPHPCDVDEGVTKTPTSQSHFYQMTPP
```

```
mRNA
/locus_tag="RP23-5084.3-004"
complement(join(17871..18150,18848..19148))
/locus_tag="RP23-5084.3-004"
/locus_tag="RP23-5084.3-004"
/product="novel protein"
/note="match: ESTs: Em:CD563843.1"
ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 169405;
Best Local Similarity 92.0%; Pred. No. 14;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAACACCACTGAGCTGGG 25
Db 159735 AAAAAACACCACTGAGCTGGG 159759
RESULT 6
AC145937/c AC145937 182785 bp DNA linear PRI 27-APR-2005
LOCUS Pan troglodytes BAC clone RP43-10P16 from chromosome 7, complete
DEFINITION sequence.
ACCESSION AC145937
VERSION AC145937.2 GI:36016669
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Pan.
1 (bases 1 to 182785)
Tomlinson,C., Cotton,M. and Kozlowicz,A.
The sequence of Pan troglodytes BAC clone RP43-10P16
Unpublished (2001)
2 (bases 1 to 182785)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 182785)
Wilson,R.
Direct Submission
Submitted (26-SEP-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 182785)
```

**AUTHORS**  
**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Wilson, R.K.  
 Direct Submission  
 Submitted (16-NOV-2004) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 5 (bases 1 to 182785)  
 Wilson, R.K.  
 Direct Submission  
 Submitted (27-APR-2005) Washington University School of Medicine, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 On Sep 26, 2003 this sequence version replaced gi:33386887.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 ----- Summary Statistics  
 Center project name: C\_PT010P16

**NOTICE:**  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
 Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

**SOURCE INFORMATION:**  
 The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clinton', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (<http://www.resgen.com>) or Pieter de Jong and co-workers at <http://www.bacpac.chori.org>.

This sequence is the entire insert of the clone.  
 Location/Qualifiers  
 1. .182785  
   /organism="Pan troglodytes"  
   /mol\_type="genomic DNA"  
   /db\_xref="taxon:9598"  
   /chromosomes="7"  
   /clone\_lib="RPCI-43"  
   /clone="RP43-10P16"  
   14981..15169  
   /notes="Unresolved simple sequence repeat."  
   26038..26707  
   /notes="Unresolved tandem repeat."

**FEATURES**  
 source  
 Query Match 87.2%; Score 21.8; DB 8; Length 182785;  
 Best Local Similarity 92.0%; Pred. No. 14;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**ORIGIN**  
 1 AAAAAACACCCACTGAGCTGGG 25  
 |||||  
 137382 AAAAAACACCCACTCATCTGGG 137358  
 |||||

**RESULT 7**  
**AC055773**  
**LOCUS**  
**DEFINITION**  
 Mus musculus clone RP23-186B19, WORKING DRAFT SEQUENCE, 3 unordered pieces.

**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

AC055773  
 AC055773.4 GI:28933881  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 214993)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-186B19  
 Unpublished  
 2 (bases 1 to 214993)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, I., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Pisani, C., Pollara, V., Oliver, J., Peterson, K., Pierre, N., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vasillef, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 214993)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, P., Cooke, P., Corum, B., DeArellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Keils, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vasillef, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 13, 2003 this sequence version replaced gi:13123353.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIER  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L5031  
 Center clone name: 186\_B\_19



```
----- Summary Statistics
Sequencing vector: M13; M77815; 37% of reads
Sequencing vector: Plasmid; n/a; 63% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 214575 bases at least Q40
Consensus quality: 214693 bases at least Q30
Consensus quality: 214775 bases at least Q20
Insert size: 220000; agarose-fp
Quality coverage: 12.1 in Q20 bases; agarose-fp
Quality coverage: 12.4 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 67177: contig of 67177 bp in length
* 67178 67277: gap of 100 bp
* 67278 91121: contig of 23844 bp in length
* 91121 91221: gap of 100 bp
* 91222 214993: contig of 123772 bp in length.

FEATURES
source
1..214993
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-186B19"
/clone_lib="RPCI-23 Female Mouse BAC"
1..67177
/note="assembly_fragment"
clone_end:SP6
vector_side:left
67178..67277
/estimated_length=100
67278..91121
/note="assembly_fragment"
91122..91221
/estimated_length=100
91222..214993
/note="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 214993;
Best Local Similarity 92.0%; Pred. No. 14;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACTGAGCTGGG 25
||||| ||||| ||||| |||||
Db 65299 AAAAAAAATCACCCTGAGCTGGG 65323

RESULT 8
AC112380/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-24L19, *** SEQUENCING IN PROGRESS
HTG 22-SBP-2002
***, 2 unordered pieces.
AC112380
AC112380.3 GI:23195384
VERSION
HTG; HTGS PHAS1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 276234)
REFERENCE
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Greggorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munitasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokeme,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villarana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 276234)
Worley,K.C.
Direct Submission
Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 276234)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21744211.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
```



```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GORW
Center clone name: CH230-24L19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 211787 bases at least Q40
Consensus quality: 214627 bases at least Q30
Consensus quality: 216603 bases at least Q20
Estimated insert size: 231080; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 15842: contig of 15842 bp in length
* 15843 15942: gap of unknown length
* 15943 276234: contig of 260292 bp in length.
FEATURES             source
1..276234
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-24L19"
4375..4751
    /note="clone_boundary"
    clone_end:5p6
    site:EcoRI
end_sequence:BH281496"
15843..15942
    /estimated_length=unknown
15943..20071
    /note="wgs contig"
complement(259013..259418)
    /note="clone_boundary"
    clone_end:T7
    site:EcoRI
end_sequence:BH281495"

ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 276234;
Best Local Similarity 92.0%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAACACACCACTGAGCTGGG 25
   |||||
Db 5299 AAGAAACACCACTGAGCTGGG 5275
   |||||

RESULT 9
AC149756
LOCUS
DEFINITION
AC149756 Bos taurus 182955 bp DNA linear MAM 31-JUL-2004
INSTITUTE Bovine BAC Library (Children's Hospital Oakland Research
INSTITUTE Bovine BAC Library (male)) complete sequence.
ACCESSION
VERSION AC149756.4 GI:50872227
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 182955)
Muzny,D,Marie, Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

```

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Bissalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,M.O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Georegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.I., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louisaged,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarpunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Steed,A., Strong,R., Sutton,A., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Thomas,S., Tingey,A., Tabor,P., Taylor,T., Taylor,T., Thomas,N., Thomas,S., Valdes,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wisczyk,R., Woodden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 182955)  
Worley,K.C.  
Direct Submission  
Submitted (19-JUN-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 182955)  
Worley,K.C.  
Direct Submission  
Submitted (28-JUL-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 182955)  
Worley,K.C.  
Direct Submission  
Submitted (31-JUL-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
COMMENT  
On Jul 31, 2004 this sequence version replaced gi:50428583.  
Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.  
FEATURES  
1. .182955  
Location/Qualifiers  
source



Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.C., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932  
2 (bases 1 to 2526)  
Klein,S. and Gerhard,D.S.  
Direct Submission  
Submitted (08-DEC-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA  
NIH-MGC Project  
Contact: XGC help desk  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Robert M. Grainger  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca

Martin Hirst, Thomas Zeng, Ryan Morin, Michelle Moksa, Johnson Pang, Diana Mah, Jing Wang, Kiech Fichter, Eric Chuah, Allen Delaney, Rob Kirkpatrick, Agnes Baross, Sarah Barber, Mabel Brown-John, Steve S. Chand, William Chow, Ryan Babakaiff, Dave Wong, Corey Matsuo, Jaclyn Beland, Susan Gibson, Luis delRio, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Mike R. Mayo, Josh Moran, Diana Palmquist, JR Santos, Duane Smalus, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Steven Jones, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 142 Row: f Column: 24  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 45361358.  
Location/Qualifiers  
1. .2526  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="MGC:76140 IMAGE:5335186"  
/tissue\_type="Embryo, Silurana tropicalis, tailbud (stages 20-27)"  
/clone\_lib="NICHD\_XGC\_Emb7"  
/lab\_host="DH10B"  
/notes="Vector: pCMV-SPORT6.1"  
1. .2526  
/gene="MGC76140"  
/db\_xref="GeneID:394869"  
61. .1056  
/gene="MGC76140"  
/codon\_start=1  
/product="hypothetical protein LOC394869"  
/protein\_id="AAH63914.1"  
/db\_xref="GI:39645383"  
/db\_xref="GeneID:394869"  
/translation="MANFNKGPAYGLSAEVKNKIAQYDPQVEELRLWEEVTGMII  
GENFOGLRDGIILCNILNKLQPGSIKKTNEAKLNWKLGNFKAKQDYGMPKPHD  
IFPANDLFENGNTQVTSILVSLAGLAKTRGFTSDVIGKYAEKQKQFGDEKWKAG  
QSVIGLQMGNTKCAQAGMTAYGTRRLHYDPKQDTKPDQDTIISLQMGNTKCAQAG  
MPAPGTRDIYDQKAILQPDVNSTSLQMGNTKCAQAGSVGLGQVYDPRYKCAAP  
TEPIIHNGSQGTGTNGSEISDSQYAEYDPQVEYDQYPRDYHGQYSDQGDIDY"

CONSRMT  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

Query Match  
Best Local Similarity 80.8%; Score 20.2; DB 5; Length 2526;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAACCAACCACTGAGCTGGG 25  
|||||  
Db 1741 AAAAAACCAACCACTGAGCTGGG 1717  
|||||

BC075515 2588 bp mRNA linear VRT 23-AUG-2004  
Xenopus tropicalis hypothetical protein MGC76140, mRNA (cDNA clone MGC:89401 IMAGE:6989687), complete cds.  
BC075515  
BC075515.1 GI:49522471  
MGC.  
Xenopus tropicalis (Silurana tropicalis)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 2588)  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,J., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.C., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932  
2 (bases 1 to 2588)  
Klein,S. and Gerhard,D.S.  
Direct Submission  
Submitted (28-JUN-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA  
NIH-MGC Project  
Contact: XGC help desk  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Robert M. Grainger  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Anura Maeson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalus, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 170 Row: g Column: 21

FEATURES  
Source

gene  
CDS

ORIGIN



```
misc_feature
gene
CDS
misc_feature
ORIGIN
Query Match      80.8%; Score 20.2; DB 8; Length 114676;
Best Local Similarity 88.0%; Pred. NO. 90;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAACACACACACCTGAGCTGGG 25
DB      33952 AAAAAACACACCAACTAGCTGGG 33976

RESULT 13
AC098690 LOCUS Homo sapiens chromosome 1 clone RP11-507I14, complete sequence.
DEFINITION AC098690
ACCESSION AC098690.3 GI:33413356
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 128000)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
Haugen, E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128000)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 128000)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE 4 (bases 1 to 128000)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2003) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Aug 2, 2003 this sequence version replaced gi:20330880.
----- Genome Center
Center: University of Washington Genome Center
Center code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
```

```
/gene="RP11-539G11.1"
/locus_tag="RP11-539G11.1-002"
Join(Complement(AC098691.2:8809..8899),
Complement(102730..102821),Complement(101148..101254),
Complement(101005..101052),Complement(97348..97445),
Complement(94446..94531),Complement(94156..94239),
Complement(93870..94027),Complement(92012..92115),
Complement(91404..91550),Complement(55037..55117),
Complement(54683..54798),Complement(52785..52864),
Complement(51767..51870),Complement(45800..45910),
Complement(44072..44162),Complement(40536..40647),
Complement(40241..40447))
/gene="RP11-539G11.1"
/locus_tag="RP11-539G11.1-002"
/product="novel protein (FLJ36760)"
/note="match: cDNAs: Em:BX649037.1"
Join(Complement(AC098691.2:8809..8839),
Complement(102730..102821),Complement(101148..101254),
Complement(101005..101052),Complement(97348..97445),
Complement(94446..94531),Complement(94156..94239),
Complement(93870..94027),Complement(92012..92115),
Complement(91404..91550),Complement(55037..55117),
Complement(54683..54798),Complement(52785..52864),
Complement(51767..51870),Complement(45800..45910),
Complement(44072..44162),Complement(40536..40647),
Complement(40292..40447))
/gene="RP11-539G11.1"
/locus_tag="RP11-539G11.1-002"
/standard_name="OTTHUMP0000011786"
/codon_start=1
/product="novel protein (FLJ36760)"
/protein_id="CA115306.1"
/db_xref="GI:55962285"
/db_xref="GOA:Q5RI13"
/db_xref="InterPro:IPR004179"
/db_xref="UniProt/TREMBL:Q5RI13"
/translaton="MDEGVNFKPTAGRLMAYIITFTVKKFYITSGKETLSLDLVTL
IAGCKEFLDIQRINEKKTILNTLNKDPNRIIRFPMEGRIKTRKMNKVLQAOAGCI
PIQDFALTQDTAKIIFHSGRIITRCITLSNAIVAGLTSFKIETDARELEILNRHP
PFQTQITWYLPKYLVEQITRYSDTTAVILTVLRNFEQLQTRTASDSHYT
LIGDADNOVVYVHLKITSVLKAGSWAKIIVKALKSEDLISLISSEFVGLDIOQ
KITVYLRPFRGNQITWQKSETQIISHKSDISTIAGPNKGTASKKPGNRCNHL
CKSKTCHGDCCKIQAOKSEIKESTISSYSLDLNRNAVSVPPVKRLKIOMNKSQV
VLKGFPTPKPSLSRSEVINSELPIEOMQDQPEIYGVKROPEYQKQKVELYN
NPELGNVWDDFDDLEVTFSFTDKTISGFGNTLSSSTRGSKLPQESKFPQR
ENSNFVSHSEMSDLSNSAMPKFSASMTKLPOQAGNAVIVHFQERKPQNLSPFEI
KQCFTFSEKNPNSSNYKKVDFPIRINSECKEYDPR"
Join(Complement(AC098691.2:8809..8839),
Complement(102730..102821),Complement(101148..101254),
Complement(101005..101052),Complement(97348..97445),
Complement(94156..94239),Complement(93870..94027),
Complement(92012..92115),Complement(91404..91550),
Complement(55037..55117),Complement(54683..54798),
Complement(52785..52864),Complement(51767..51870),
Complement(45800..45910),Complement(40571..40647))
/gene="RP11-539G11.1"
/locus_tag="RP11-539G11.1-002"
/standard_name="OTTHUMP0000011785"
/note="match: proteins: Tr:Q8N9Q0"
/codon_start=1
/product="novel protein (FLJ36760)"
/protein_id="CA115307.1"
/db_xref="GI:55962286"
/db_xref="GOA:Q8N9Q0"
/db_xref="InterPro:IPR004179"
/db_xref="UniProt/TREMBL:Q8N9Q0"
/translaton="MDEGVNFKPTAGRLMAYIITFTVKKFYITSGKETLSLDLVTL
IAGCKEFLDIQRINEKKTILNTLNKDPNRIIRFPMEGRIKTRKMNKVLQAOAGCI
PIQDFALTQDTAKIIFHSGRIITRWLSDFVAAQKFPVALLSLILAKCRCKWENSL
HYVKOLEKIGITLSNAIVAGLTSFKIETDARELEILNRHPPTQITKETVMYLP
KYELKVEQITRYSDTAETLVILNFEQLQTRTASDSHYVTLIGDADNOVVYVHL
KITDSVLLKAGSWAKIIVKALKSEDLISLISSEFVGLDIOQKITVYLRPFRGN
```

----- Project Information  
Center project name: chr-1  
Center clone name: RP11-507114 (sc0371)  
----- Summary Statistics  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator ET; 49% of reads  
Assembly: Dye-terminator Big Dye; 51% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 127991 bases at least Q40  
Consensus quality: 127997 bases at least Q30  
Consensus quality: 128000 bases at least Q20  
Insert size: 128000; sum-of-contigs  
Quality coverage: 19.0x in Q20 bases; sum-of-contigs  
-----  
Overlapping Sequences:  
5': RF4-648113 (UWGC:sc0196) AC093563, 110680-bp overlap  
3': RP11-539G11, BX323048, 45716-bp overlap  
Note: This is a partial submission. The full clone overlaps are  
not included.

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:

all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII				EcoRI				HindIII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
3177	3147	8696	8841	4087	4011						
2067	2090	6	<800	6382	6489						
11746	11739	2995	3061	512	<800						
9669	9454	5942	6031	449	<800						
1898	1917	3870	3904	1289	1263						
450	<800	2692	2709	14	<800						
1732	1742	1474	1448	1355	1327						
1038	1020	4248	4255	1943	1915						
1291	1271	1160	1146	6484	6489						
7167	7290	10782	10713	1793	1915						

4631	4627	1448	1448	1322	1327
1533	1524	5710	5687	7207	7323
1035	1020	3032	3061	12	<800
2039	2090	2692	2709	316	<800
6302	6289	10021	10083	5438	5542
1276	1271	941	942	7483	7323
10795	10817	4920	4957	361	<800
1079	1020	3375	3439	1133	1138
206	<800	9374	9285	1740	1773
4633	4627	2974	2909	181	<800
964	1020	4638	4637	2549	2554
478	<800	6110	6031	7689	7323
680	<800	3061	3061	1548	1556
1433	1424	313	<800	3231	3273
5765	5740	2934	2909	5632	5542
6590	6607	1809	1868	59	<800
9335	9454	5126	4957	354	<800
792	<800	1849	1868	1132	1138
5538	5483	11502	11321	274	<800
3107	3147	9616	9605	1787	1773
973	1020	3438	3439	1643	1616
2139	2090	3061	3061	2023	2112
8116	8094	197	<800	1280	1263
748	<800	13507	13631	338	<800
1910	1917	1893	1868	48	<800
2159	2090	4888	4957	578	<800
2441	2471	354	<800	5807	5542
1938	1917	3986	3904	1891	1915
778	<800			7303	7323
6140	6289			503	<800
2505	2471			100	<800
7054	6985			1419	1416
1435	1424			122	<800
4273	4244			8785	8814
6253	6289			1533	1556
1000	1020			154	<800
2046	2090			48	<800

```
-----
4280      4244      2150      2112
-----
1078      1069
-----
359       <800
-----
24        <800
-----
2531      2554
-----
244       <800
-----
8018      7990

Query Match      80.8%; Score 20.2; DB 8; Length 128000;
Best Local Similarity 88.0%; Pred. No. 90;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 116235 AAAAAACCAACCAATAGCTGGG 116259

RESULT 14
AC015968
LOCUS      128638 bp      DNA      linear      PRI 15-OCT-2003
DEFINITION Homo sapiens BAC clone RP11-133L20 from 7, complete sequence.
ACCESSION  AC015968
VERSION     AC015968.4 GI:10716639
KEYWORDS   HTG
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 128638)
AUTHORS   Sulston,J.E. and Wilson,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
PUBMED    9847074
REFERENCE  2 (bases 1 to 128638)
AUTHORS   Joshi,C., Kistic,M., Laplant,Y. and Maupin,R.
TITLE     The sequence of Homo sapiens BAC clone RP11-133L20
JOURNAL   Unpublished (2001)
REFERENCE  3 (bases 1 to 128638)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (17-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  4 (bases 1 to 128638)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (07-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE  5 (bases 1 to 128638)
AUTHORS   Wilson,R.
TITLE     Direct Submission
JOURNAL   Submitted (15-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Oct 7, 2000 this sequence version replaced gi:7630858.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_NH0133L20
            -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-457C23 the clone sequenced to the right is RP5-953F6, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-133L20 actual end is at base position 37782 of RP5-953F6.

#### FEATURES

	Location/Qualifiers	source
	1..128638	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7"	
repeat_region	356..394	
	/rpt_family="TATG)n"	
repeat_region	402..663	
	/rpt_family="Alu"	
repeat_region	991..1295	
	/rpt_family="Alu"	
repeat_region	2025..2055	
	/rpt_family="(TA)n"	
repeat_region	2057..2088	
	/rpt_family="(GA)n"	
repeat_region	2089..2363	
	/rpt_family="Alu"	
repeat_region	2441..2542	
	/rpt_family="MIR"	
repeat_region	2569..3173	
	/rpt_family="ERV1"	
repeat_region	5618..5926	
	/rpt_family="Alu"	
repeat_region	5967..6231	
	/rpt_family="Alu"	
repeat_region	6418..6949	
	/rpt_family="MER1_type"	
repeat_region	7921..8227	
	/rpt_family="Alu"	
repeat_region	8398..8478	
	/rpt_family="MIR"	
repeat_region	8558..8696	
	/rpt_family="MIR"	



```

repeat_region 11213..11247
repeat_region /rpt_family="Aln"
repeat_region 14788..14814
repeat_region /rpt_family="(TTA)n"
repeat_region 14815..15576
repeat_region /rpt_family="L1"
repeat_region 15574..16532
repeat_region /rpt_family="L1"
repeat_region 16762..17064
repeat_region /rpt_family="Alu"
repeat_region 20843..22355
repeat_region /rpt_family="L1"
repeat_region 22359..22696
repeat_region /rpt_family="L1"
repeat_region 22706..23010
repeat_region /rpt_family="Alu"
repeat_region 23834..24127
repeat_region /rpt_family="Alu"
repeat_region 24154..24601
repeat_region /rpt_family="L1"
repeat_region 25597..26090
repeat_region /rpt_family="L1"
repeat_region 26229..26249
repeat_region /rpt_family="AT-rich"
repeat_region 29072..29257
repeat_region /rpt_family="L1"
repeat_region 29450..29851
repeat_region /rpt_family="L1"
repeat_region 29841..30049
repeat_region /rpt_family="L1"
repeat_region 30050..30121
repeat_region /rpt_family="Alu"
repeat_region 30122..30693
repeat_region /rpt_family="L1"
repeat_region 30694..30996
repeat_region /rpt_family="Alu"
repeat_region 30997..31396
repeat_region /rpt_family="L1"
repeat_region 31397..31700
repeat_region /rpt_family="Alu"
repeat_region 31701..32174
repeat_region /rpt_family="L1"
repeat_region 32198..32500
repeat_region /rpt_family="Alu"
repeat_region 32502..32924
repeat_region /rpt_family="MaLR"
repeat_region 32939..32994
repeat_region /rpt_family="A-rich"
repeat_region 34831..35126
repeat_region /rpt_family="Alu"
repeat_region 35193..35649
repeat_region /rpt_family="L1"
repeat_region 35863..36526
repeat_region /rpt_family="ERV1"
repeat_region 36527..36951
repeat_region /rpt_family="ERV1"
repeat_region 37518..37562
repeat_region /rpt_family="AT-rich"
repeat_region 37630..37897
repeat_region /rpt_family="L1"
repeat_region 38642..38967
repeat_region /rpt_family="ERV1"
repeat_region 39307..39606
repeat_region /rpt_family="Alu"
repeat_region 40032..40362
repeat_region /rpt_family="MaLR"
repeat_region 40987..41230
repeat_region /rpt_family="L1"
repeat_region 41658..41954
repeat_region /rpt_family="Alu"
repeat_region 42028..42085
repeat_region /rpt_family="L1"
repeat_region 42086..42361

```

```

/rpt_family="L1"
42465..43084
/rpt_family="L1"
43084..44825
/rpt_family="L1"
43512..43698
/note="unresolved homopolymeric repeat"

Query Match      80.8%; Score 20.2; DB 8; Length 128638;
Best Local Similarity 88.0%; Pred. No. 90;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25
Db 11242 AAAAAACACACCACTGAGCTGGG 11266
|||||||
AL935194 145569 bp DNA linear VRT 13-DEC-2002
Zebrafish DNA sequence from clone CH211-237E12, complete sequence.
AL935194
ACCESSION
VERSION AL935194.4 GI:26985414
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 145569)
Giselle,H.
Direct Submission
Submitted (12-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Dec 16, 2002 this sequence version replaced gi:24940082.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zface@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
beginning 'Dr' were identified by the Recon repeat discovery system
(Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see http://www/Projects/D_rerio/fishmask.shtml
CH211-237E12 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
FEATURES
Location/Qualifiers
1..145569
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-237E12"

```

source

```
ORIGIN
/clone_lib="CHORI-211"

Query Match      80.8%; Score 20.2; DB 5; Length 145569;
Best Local Similarity 88.0%; Pred. No. 90;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAACACCACTGAGCTGGG 25
Db 33473 AAAAAACACCACTGAGCTGGG 33497

RESULT 16
AC053517/c
LOCUS AC053517.7 GI:15451722 linear HTG 06-SEP-2001
DEFINITION Homo sapiens chromosome 4 clone RP11-455G12, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION AC053517
VERSION AC053517.7 GI:15451722
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
REFERENCE 1 (bases 1 to 155263)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155263)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
COMMENT On Sep 6, 2001 this sequence version replaced gi:11192229.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0455G12
----- Summary Statistics -----
Sequencing vector: M13; 99%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 99% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150711 bases at least Q40
Consensus quality: 152938 bases at least Q30
Consensus quality: 154294 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1339: contig of 1339 bp in length
* 1340: contig of unknown length
* 1440: contig of 2429 bp in length
* 3869: gap of unknown length
* 3969: contig of 2989 bp in length
* 6958: gap of unknown length
* 7058: contig of 3693 bp in length
* 10751: gap of unknown length
* 10851: contig of 4158 bp in length
* 10851: gap of unknown length
* 15009: gap of unknown length
* 15109: contig of 5421 bp in length
*
* 20530: gap of unknown length
* 20630: contig of 4515 bp in length
* 25145: gap of unknown length
* 25245: gap of 7395 bp in length
* 32739: gap of unknown length
* 32740: contig of 8751 bp in length
* 41491: gap of unknown length
* 41591: gap of unknown length
* 56100: contig of 14510 bp in length
* 56101: gap of unknown length
* 56201: contig of 15145 bp in length
* 71346: gap of unknown length
* 90217: contig of 18771 bp in length
* 90317: gap of unknown length
* 103167: contig of 19051 bp in length
* 103668: gap of unknown length
* 109468: contig of 45796 bp in length.
          Location/Qualifiers
            1..155263
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="4"
              /clone="RP11-455G12"
            1..1339
              /note="assembly_name:Contig68"
            1340..1439
              /estimated_length=unknown
            1440..3868
              /note="assembly_name:Contig69"
              clone_end:SP6
              vector_side:right"
            3869..3968
              /estimated_length=unknown
            3969..6957
              /note="assembly_name:Contig70"
            6958..7057
              /estimated_length=unknown
            7058..10750
              /note="assembly_name:Contig71"
            10751..10850
              /estimated_length=unknown
            10851..15008
              /note="assembly_name:Contig72"
            15009..15108
              /estimated_length=unknown
            15109..20529
              /note="assembly_name:Contig73"
            20530..20629
              /estimated_length=unknown
            20630..25144
              /note="assembly_name:Contig74"
            25145..25244
              /estimated_length=unknown
            25245..32639
              /note="assembly_name:Contig75"
            32640..32739
              /estimated_length=unknown
            32740..41490
              /note="assembly_name:Contig76"
            41491..41590
              /estimated_length=unknown
            41591..56100
              /note="assembly_name:Contig77"
            56101..56200
              /estimated_length=unknown
            56201..71345
              /note="assembly_name:Contig78"
              clone_end:T7
              vector_side:right"
            71346..71445
              /estimated_length=unknown
            71446..90216
              /note="assembly_name:Contig79"
```

gap 90217. .90316  
/estimated length=unknown  
misc\_feature 90317. .109367  
/note="assembly\_name:Contig80"  
gap 109368. .109467  
/estimated length=unknown  
misc\_feature 109468. .155263  
/note="assembly\_name:Contig81"  
ORIGIN  
Query Match 80.8%; Score 20.2; DB 14; Length 155263;  
Best Local Similarity 88.0%; Pred. No. 91;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAACACACCACTGAGCTGGG 25  
Db 53040 AAAAAACACACCACTGACTAGG 53016  
RESULT 17  
AC016124/c 159976 bp DNA linear HTG 20-SEP-2000  
LOCUS Homo sapiens clone RP11-28G17, WORKING DRAFT SEQUENCE, 20 unordered  
DEFINITION pieces.  
ACCESSION AC016124  
VERSION AC016124.4 GI:10198467  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 159976)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
TITLE Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
JOURNAL Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
REFERENCE 2 (bases 1 to 159976)  
AUTHORS Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Doyle,M.,  
TITLE Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
JOURNAL Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
REFERENCE Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
AUTHORS Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
TITLE Lebecky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
JOURNAL McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
REFERENCE Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
AUTHORS Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
TITLE Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
JOURNAL Teefave,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
REFERENCE Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome  
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 20, 2000 this sequence version replaced gi:9137133.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4756  
Center clone name: 28 G 17  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 150281 bases at least Q40  
Consensus quality: 154839 bases at least Q30  
Consensus quality: 156664 bases at least Q20

Insert size: 175000; agarose-fp  
Insert size: 158076; sum-of-contigs  
Quality coverage: 4.4 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1748: contig of 1748 bp in length  
\* 1749: gap of 100 bp  
\* 3313: contig of 1465 bp in length  
\* 3314: gap of 100 bp  
\* 4508: contig of 1095 bp in length  
\* 4608: gap of 100 bp  
\* 6383: contig of 1775 bp in length  
\* 6384: gap of 100 bp  
\* 8884: contig of 2401 bp in length  
\* 8885: gap of 100 bp  
\* 11538: contig of 2554 bp in length  
\* 11539: gap of 100 bp  
\* 15115: contig of 3477 bp in length  
\* 15116: gap of 100 bp  
\* 18012: contig of 2797 bp in length  
\* 18013: gap of 100 bp  
\* 22355: contig of 4243 bp in length  
\* 22356: gap of 100 bp  
\* 22456: contig of 4105 bp in length  
\* 26560: gap of 100 bp  
\* 26561: contig of 4135 bp in length  
\* 30795: gap of 100 bp  
\* 30796: contig of 4360 bp in length  
\* 35255: gap of 100 bp  
\* 35256: contig of 6388 bp in length  
\* 41743: gap of 100 bp  
\* 41744: contig of 40791 bp in length  
\* 82634: gap of 100 bp  
\* 82734: gap of 100 bp  
\* 89758: contig of 7024 bp in length  
\* 89759: gap of 100 bp  
\* 100523: contig of 10665 bp in length  
\* 100524: gap of 100 bp  
\* 111515: contig of 10892 bp in length  
\* 111516: gap of 100 bp  
\* 132549: contig of 20934 bp in length  
\* 132550: gap of 100 bp  
\* 132550: contig of 23020 bp in length  
\* 155670: gap of 100 bp  
\* 155670: contig of 4207 bp in length.  
\* 155770: Location/Qualifiers  
1. 159976  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-28G17"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 1748  
/note="assembly\_fragment"  
clone end:SP6  
vector\_side:left"  
1749. 1848  
/estimated\_length=100  
1849. 3313  
/note="assembly\_fragment"  
3314. 3413  
/estimated\_length=100  
3414. 4508  
/note="assembly\_fragment"  
4509. 4608  
/estimated\_length=100

misc\_feature 4609. .6383 /note="assembly\_fragment"  
gap 6384. .6483 /estimated\_length=100  
misc\_feature 6484. .8884 /note="assembly\_fragment"  
gap 8885. .8984 /estimated\_length=100  
misc\_feature 8985. .11538 /note="assembly\_fragment"  
gap 11539. .11638 /estimated\_length=100  
misc\_feature 11639. .15115 /note="assembly\_fragment"  
gap 15116. .15215 /estimated\_length=100  
misc\_feature 15216. .18012 /note="assembly\_fragment"  
gap 18013. .18112 /estimated\_length=100  
misc\_feature 18113. .22355 /note="assembly\_fragment"  
gap 22356. .22455 /estimated\_length=100  
misc\_feature 22456. .26560 /note="assembly\_fragment"  
gap 26561. .26660 /estimated\_length=100  
misc\_feature 26661. .30795 /note="assembly\_fragment"  
gap 30796. .30895 /estimated\_length=100  
misc\_feature 30896. .35255 /note="assembly\_fragment"  
gap 35256. .35355 /estimated\_length=100  
misc\_feature 35356. .41743 /note="assembly\_fragment"  
gap 41744. .41843 /estimated\_length=100  
misc\_feature 41844. .82634 /note="assembly\_fragment"  
gap 82635. .82734 /estimated\_length=100  
misc\_feature 82735. .89758 /note="assembly\_fragment"  
gap 89759. .89858 /estimated\_length=100  
misc\_feature 89859. .100523 /note="assembly\_fragment"  
gap 100524. .100623 /estimated\_length=100  
misc\_feature 100624. .111515 /note="assembly\_fragment"  
gap 111516. .111615 /estimated\_length=100  
misc\_feature 111616. .132549 /note="assembly\_fragment"  
gap 132550. .132649 /estimated\_length=100  
misc\_feature 132650. .155669 /note="assembly\_fragment"  
gap 155670. .155769 /estimated\_length=100  
misc\_feature 155770. .159976 /note="assembly\_fragment"  
clone\_end:77  
vector\_side:right"

ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 159976;  
Best Local Similarity 88.0%; Pred. No. 91;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  
Db  
RESULT 18  
AC073301  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 AAAAAACACCACTGAGCTGGG 25  
|||||  
111399 AAAAAACACCACTGATCTAGG 111375  
|||||  
AC073301 173256 bp DNA linear HTG 10-MAY-2001  
Homo sapiens chromosome 1 clone RP11-221N16 map 1, WORKING DRAFT  
SEQUENCE, 23 unordered pieces.  
AC073301  
AC073301.2 GI:14017530  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 173256)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 1, clone RP11-221N16  
Unpublished  
2 (bases 1 to 173256)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Collins,S.,  
Compiano,A., Castle,A., Chosepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (14-JUN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 10, 2001 this sequence version replaced gi:8516089.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7730  
Center clone name: 221\_N\_16  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 160406 bases at least Q40  
Consensus quality: 166935 bases at least Q30  
Consensus quality: 169615 bases at least Q20  
Insert size: 163000; agarose-fp  
Insert size: 171056; sum-of-contigs  
Quality coverage: 4.3 in Q20 bases; agarose-fp  
Quality coverage: 4.1 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 7400: contig of 7400 bp in length  
\* 7401 7500: gap of 100 bp  
\* 7501 8604: contig of 1104 bp in length  
\* 8605 8705: gap of 100 bp  
\* 29589: contig of 20885 bp in length  
\* 29590 31266: contig of 1577 bp in length  
\* 31267 31366: gap of 100 bp  
\* 31367 32511: contig of 1145 bp in length  
\* 32512 32611: gap of 100 bp  
\* 32612 35694: contig of 3083 bp in length  
\* 35695 35794: gap of 100 bp  
\* 35795 39437: contig of 3643 bp in length  
\* 39438 39537: gap of 100 bp  
\* 39538 43511: contig of 3974 bp in length  
\* 43512 43611: gap of 100 bp  
\* 43612 47479: contig of 3868 bp in length  
\* 47480 47579: gap of 100 bp  
\* 47580 53487: contig of 5908 bp in length  
\* 53488 53587: gap of 100 bp  
\* 53588 58391: contig of 4804 bp in length  
\* 58392 64453: contig of 5962 bp in length  
\* 64454 64553: gap of 100 bp  
\* 64554 70748: contig of 6195 bp in length  
\* 70749 70848: gap of 100 bp  
\* 70849 76428: contig of 5580 bp in length  
\* 76429 85292: contig of 8764 bp in length  
\* 85293 91957: contig of 100 bp  
\* 91958 92057: gap of 100 bp  
\* 92058 101108: contig of 9051 bp in length  
\* 101109 101208: gap of 100 bp  
\* 101209 111112: contig of 9904 bp in length  
\* 111113 121129: gap of 100 bp  
\* 121130 121229: contig of 9917 bp in length  
\* 121230 134712: gap of 100 bp  
\* 134713 134812: contig of 13483 bp in length  
\* 134813 152688: contig of 17876 bp in length  
\* 152689 152788: gap of 100 bp  
\* 152789 172184: contig of 19396 bp in length  
\* 172185 172284: gap of 100 bp  
\* 172285 173256: contig of 972 bp in length.

## FEATURES

Source  
1..173256  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"  
/clone="RP11-221N16"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature  
1..7400  
/note="assembly\_fragment  
clone\_end:SP6  
vector side:left"  
7401..7500  
/estimated\_length=100  
misc\_feature  
7501..8604  
/note="assembly\_fragment"  
8605..8704  
/estimated\_length=100  
misc\_feature  
8705..29589  
/note="assembly\_fragment"

gap  
misc\_feature  
29590..29589  
/estimated\_length=100  
misc\_feature  
29690..31266  
/note="assembly\_fragment"  
31267..31366  
/estimated\_length=100  
misc\_feature  
31367..32511  
/note="assembly\_fragment"  
32512..32611  
/estimated\_length=100  
misc\_feature  
32612..35694  
/note="assembly\_fragment"  
35695..35794  
/estimated\_length=100  
misc\_feature  
35795..39437  
/note="assembly\_fragment"  
39438..39537  
/estimated\_length=100  
misc\_feature  
39538..43511  
/note="assembly\_fragment"  
43512..43611  
/estimated\_length=100  
misc\_feature  
43612..47479  
/note="assembly\_fragment"  
47480..47579  
/estimated\_length=100  
misc\_feature  
47580..53487  
/note="assembly\_fragment"  
53488..53587  
/estimated\_length=100  
misc\_feature  
53588..58391  
/note="assembly\_fragment"  
58392..58491  
/estimated\_length=100  
misc\_feature  
58492..64453  
/note="assembly\_fragment"  
64454..64553  
/estimated\_length=100  
misc\_feature  
64554..70748  
/note="assembly\_fragment"  
70749..70848  
/estimated\_length=100  
misc\_feature  
70849..76428  
/note="assembly\_fragment"  
76429..76528  
/estimated\_length=100  
misc\_feature  
76529..85292  
/note="assembly\_fragment"  
85293..85392  
/estimated\_length=100  
misc\_feature  
85393..91957  
/note="assembly\_fragment"  
91958..92057  
/estimated\_length=100  
misc\_feature  
92058..101108  
/note="assembly\_fragment"  
101109..101208  
/estimated\_length=100  
misc\_feature  
101209..111112  
/note="assembly\_fragment"  
11113..111212  
/estimated\_length=100  
gap  
misc\_feature  
164289  
/estimated\_length=100

Query Match 80.8%; Score 20.2; DB 14; Length 173256;  
Best Local Similarity 88.0%; Pred. No. 91;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCCACCTAGCTGGG 25  
|||||  
Db 164289 AAAAAACACCCACCTAGCTGGG 164313  
|||||

RESULT 19

AC013557/c  
LOCUS  
DEFINITION  
AC013557  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC013557 178806 bp DNA linear HTG 24-AUG-2002  
Homo sapiens chromosome 11 clone RP11-28D2 map 11, WORKING DRAFT  
SEQUENCE, 4 unordered pieces.

AC013557.3 GI:7107807  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 178806)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 11, clone RP11-28D2

Unpublished

2 (bases 1 to 178806)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,

Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 178806)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,

Camataca,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,

Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,

Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,

Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K.,

Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,

Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,

McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,

Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,

Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,

Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 28, 2000 this sequence version replaced gi:6479117.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L2252

Center clone name: 28 D 2

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 170392 bases at least Q40  
Consensus quality: 176132 bases at least Q30  
Consensus quality: 177464 bases at least Q20  
Insert size: 176000; agarose-fp  
Insert size: 178506; sum-of-contigs  
Quality coverage: 6.7 in Q20 bases; agarose-fp  
Quality coverage: 6.6 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1239: contig of 1239 bp in length  
\* 1240 1339: gap of 100 bp  
\* 1340 23256: contig of 21917 bp in length  
\* 23257 23356: gap of 100 bp  
\* 23357 88091: contig of 64735 bp in length  
\* 88092 88191: gap of 100 bp  
\* 88192 178806: contig of 90615 bp in length.

#### FEATURES

source

1..178806

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="RP11-28D2"

/clone\_lib="RPC1-11 Human Male BAC"

1..1239

/note="assembly\_fragment"

misc\_feature

gap

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

## REFERENCE

1 (bases 1 to 181489)

## AUTHORS

Sulston, J.E. and Waterston, R.

## TITLE

Toward a complete human genome sequence

## JOURNAL

Genome Res. 8 (11), 1097-1108 (1998)

## PUBMED

9847074

## REFERENCE

2 (bases 1 to 181489)

## AUTHORS

Scott, K., Haglund, K. and Doebber, A.

## TITLE

The sequence of Homo sapiens BAC clone RP11-804N11

## JOURNAL

Unpublished (2001)

## REFERENCE

3 (bases 1 to 181489)

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## REFERENCE

4 (bases 1 to 181489)

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (03-MAY-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## REFERENCE

6 (bases 1 to 181489)

## AUTHORS

Waterston, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (29-MAY-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

## COMMENT

On May 3, 2002 this sequence version replaced gi:18693591.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)

----- Summary Statistics

-----

Center project name: H\_NH0804N11

-----

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this

sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male

donor, as described by Oseegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from

Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong

and coworkers at <http://www.chori.org>

VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-703H20; the clone sequenced

to the right is RP11-556N4. Actual start of this clone is at base

position 1995 of RP11-703H20; actual end is at base position 181489

of RP11-804N11.

The region from 97601 to 97728 is covered only by a per product

from clone DNA. Data from AC053517 was used to finish AC104828.

## FEATURES

## Source

1. 181489

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-804N11"

/clone\_lib="RP11-11"

7. 107

/rpt\_family="L1"

107. 405

/rpt\_family="L1"

406. 713

/rpt\_family="Alu"

714. 906

/rpt\_family="L1"

946. 1295

/rpt\_family="Alu"

1543. 1844

/rpt\_family="Alu"

2424. 2718

/rpt\_family="Alu"

4417. 4456

/rpt\_family="AT\_rich"

4975. 5000

/rpt\_family="(CA)n"

5184. 5213

/rpt\_family="AT\_rich"

5276. 5572

/rpt\_family="Alu"

5774. 6138

/rpt\_family="L2"

6271. 6296

/rpt\_family="AT\_rich"

6303. 6624

/rpt\_family="(CA)n"

6759. 7055

/rpt\_family="Alu"

7036. 7156

/rpt\_family="(GGAA)n"

8574. 8600

/rpt\_family="(CA)n"

8977. 9022

/rpt\_family="L2"

9131. 9160

/rpt\_family="AT\_rich"

9327. 9350

/rpt\_family="AT\_rich"

9463. 10389

/rpt\_family="MER1\_type"

10390. 10684

/rpt\_family="Alu"

10685. 10860

/rpt\_family="MER1\_type"

11528. 11556

/rpt\_family="AT\_rich"

11575. 11868

/rpt\_family="Alu"

11918. 12357

/rpt\_family="L1"

12434. 12458

/rpt\_family="(TTTTA)n"

12566. 12755

/rpt\_family="L1"

repeat\_region



```

repeat_region 13279 .13587
/rpt family="Alu"
repeat_region 13630 .13931
/rpt family="Alu"
repeat_region 14094 .14409
/rpt family="Alu"
repeat_region 14559 .14838
/rpt family="Alu"
repeat_region 15975 .16123
/rpt family="Alu"
repeat_region 16182 .16466
/rpt family="Alu"
repeat_region 16467 .16491
/rpt family="CA)n"
repeat_region 16768 .16989
/rpt family="MER2_type"
repeat_region 17327 .17622
/rpt family="Alu"
repeat_region 18500 .18604
/rpt family="L2"
repeat_region 18932 .19222
/rpt family="Alu"
repeat_region 19233 .19534
/rpt family="Alu"
repeat_region 19822 .20133
/rpt family="Alu"
repeat_region 20435 .20716
/rpt family="Alu"
repeat_region 20717 .20786
/rpt family="TA)n"
repeat_region 21032 .21341
/rpt family="Alu"
repeat_region 21342 .21498
/rpt family="MTR"
repeat_region 21794 .21823
/rpt family="AT_rich"
repeat_region 21918 .21962
/rpt family="TATG)n"
repeat_region 22071 .22191
/rpt family="MTR"
repeat_region 22197 .22286
/rpt family="L2"
repeat_region 22287 .22590
/rpt family="Alu"
repeat_region 22591 .22686
/rpt family="L2"

Query Match 80.8%; Score 20.2; DB 8; Length 181489;
Best Local Similarity 88.0%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACACCACTGAGCTGGG 25
Db 86906 AAAAAACACACCACTGATCTAGG 86930

RESULT 21
AC099068
LOCUS
DEFINITION Rattus norvegicus clone CH230-7F9, WORKING DRAFT SEQUENCE, 5
ACCESSION AC099068
VERSION AC099068.9 GI:30467914
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,

```

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,J., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.B., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nair,L., Nwackemehe,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Steimle,M., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,P., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 243279)  
Worley,K.C.  
Direct Submission  
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 243279)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 9, 2003 this sequence version replaced gi:24941198.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

ORGANISM	Rattus norvegicus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 256914)	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alibrooks, S., Amin, A., Angulano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. D., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto-Mo, Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 256914)	
AUTHORS	Worley, K. C.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
REFERENCE	3 (bases 1 to 256914)	
AUTHORS	Rat Genome Sequencing Consortium.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
COMMENT	On May 9, 2003 this sequence version replaced gi:24817816. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas	

----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a> ----- Project Information Center project name: GCEH Center clone name: CH230-7F9 ----- Summary Statistics Assembly program: Atlas; Consensus quality: 237282 bases at least Q40 Consensus quality: 238711 bases at least Q30 Consensus quality: 239637 bases at least Q20 Estimated insert size: 245767; sum-of-contigs estimation Quality coverage: 11x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html</a> ). * NOTE: This is a 'working draft' sequence. It currently * consists of 5 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 231160: contig of 231160 bp in length * 231161 231260: gap of unknown length * 231261 236392: contig of 5132 bp in length * 236393 236492: gap of unknown length * 236493 240099: contig of 3607 bp in length * 240100 240199: gap of unknown length * 240200 241548: contig of 1349 bp in length * 241549 241648: gap of unknown length * 241649 243279: contig of 1631 bp in length. * * FEATURES * source 1. 243279 * /organism="Rattus norvegicus" * /mol_type="genomic DNA" * /db_xref="taxon:10116" * /clone="CH230-7F9" * * misc_feature 1. 1310 * /note="wgs contig" * * gap 231161..231260 * /estimated_length=unknown * * misc_feature 231261..232384 * /note="wgs contig" * * gap 236393..236492 * /estimated_length=unknown * * gap 240100..240199 * /estimated_length=unknown * * gap 241549..241648 * /estimated_length=unknown * * ORIGIN Query Match 80.8%; Score 20.2; DB 14; Length 243279; Best Local Similarity 88.0%; Pred. No. 91; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  Qy 1 AAAAAACACACACTGAGCTGGG 25                               Db 137647 AAAAAAACACCTGTGAGCTGGG 137671  RESULT 22 AC095147/c LOCUS AC095147 256914 bp DNA linear HTG 09-MAY-2003 DEFINITION Rattus norvegicus clone CH230-8H22, WORKING DRAFT SEQUENCE, 3 unordered pieces. AC095147 AC095147.6 GI:30467274 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. KEYWORDS Rattus norvegicus (Norway rat) SOURCE
---

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GTC

Center clone name: CH230-8H22

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 205127 bases at least Q40

Consensus quality: 208686 bases at least Q30

Consensus quality: 210795 bases at least Q20

Estimated insert size: 213188; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 253811: contig of 253811 bp in length  
\* 253812 253911: gap of unknown length  
\* 253912 255439: contig of 1528 bp in length  
\* 255440 255539: gap of unknown length  
\* 255540 256914: contig of 1375 bp in length.

Location/Qualifiers

1. .256914  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-8H22"  
258. .925

misc\_feature

/note="clone boundary  
clone end:Sp6  
site:ECORI  
end\_sequence:BH345121"

complement(245885..252580)  
/note="clone boundary  
clone end:T7  
site:ECORI  
end\_sequence:BH345119"

complement(251898..252621)  
/note="clone boundary  
clone end:T7  
site:ECORI  
end\_sequence:BH345119"

253812..253911

/estimated\_length=unknown

255440..255539

/estimated\_length=unknown

ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 256914;  
Best Local Similarity 88.0%; Pred. No. 91;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25

|||||  
Db 60736 AAAAAACACCACTGAGCAAGG 60712

RESULT 23

AC160578

LOCUS

DEFINITION

AC160578

AC160578.2 GI:58300415

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Bos taurus (cow)

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

AC160578 290052 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-8A20, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 48  
unordered pieces.

AC160578  
AC160578.2 GI:58300415  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Bos taurus (cow)

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

Bos taurus

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 290052)  
Worley,K.C.

Direct Submission

Submitted (28-APR-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 290052)  
Cov Genome Sequencing Consortium.  
Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jun 29, 2005 this sequence version replaced gi:62945466.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: FFGH  
Center clone name: CH240-84A20  
----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 268871 bases at least Q40  
Consensus quality: 273631 bases at least Q30  
Consensus quality: 277456 bases at least Q20  
Estimated insert size: 274073; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 48 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 7000: contig of 7000 bp in length  
7001 7050: gap of 50 bp  
7051 9676: contig of 2626 bp in length  
9677 9776: gap of unknown length  
9777 16485: contig of 6709 bp in length  
16486 16535: gap of 50 bp  
16536 17680: contig of 1145 bp in length  
17681 22053: gap of 4373 bp  
22054 24414: contig of 2361 bp in length  
24415 24464: gap of 50 bp  
24465 30888: contig of 6424 bp in length  
30889 30939: gap of 50 bp  
40084: contig of 9146 bp in length  
40085 40134: gap of 50 bp  
40135 45676: contig of 5542 bp in length  
45677 45726: gap of 50 bp  
45727 57298: contig of 11572 bp in length  
57299 57348: gap of 50 bp  
57349 64857: contig of 7509 bp in length  
64858 64907: gap of 50 bp  
64908 70695: contig of 5788 bp in length  
70696 70795: gap of unknown length  
70796 77607: contig of 6812 bp in length  
77608 77657: gap of 50 bp  
77658 81575: contig of 3918 bp in length  
81576 81676: gap of unknown length  
81677 85723: contig of 4048 bp in length

\* 85724 85773: gap of 50 bp  
85774 93143: contig of 7370 bp in length  
93144 93193: gap of 50 bp  
93194 94931: contig of 1738 bp in length  
94932 94981: gap of 50 bp  
94982 112800: contig of 17819 bp in length  
112801 112850: gap of 50 bp  
112851 115975: contig of 3125 bp in length  
115976 116025: gap of 50 bp  
116026 121069: contig of 5044 bp in length  
121070 121119: gap of 50 bp  
121120 126466: contig of 5347 bp in length  
126467 126516: gap of 50 bp  
126517 142203: contig of 15687 bp in length  
142204 142303: gap of unknown length  
142304 147764: contig of 5461 bp in length  
147765 147814: gap of 50 bp  
147815 152013: contig of 4199 bp in length  
152014 152063: gap of 50 bp  
152064 157204: contig of 5141 bp in length  
157205 157254: gap of 50 bp  
157255 209028: contig of 51774 bp in length  
209029 209078: gap of 50 bp  
209079 212455: contig of 3377 bp in length  
212456 212995: gap of 540 bp  
212996 214790: contig of 1795 bp in length  
214791 214840: gap of 50 bp  
214841 216146: contig of 1306 bp in length  
216147 216246: gap of unknown length  
216247 226185: contig of 9939 bp in length  
226186 226235: gap of 50 bp  
226236 233347: contig of 7112 bp in length  
233348 233447: gap of unknown length  
233448 235721: contig of 2274 bp in length  
235722 235771: gap of 50 bp  
235772 237540: contig of 1769 bp in length  
237541 237640: gap of unknown length  
237641 246873: contig of 9233 bp in length  
246874 246923: gap of 50 bp  
246924 252361: contig of 5438 bp in length  
252362 252411: gap of 50 bp  
252412 255888: contig of 3477 bp in length  
255889 255938: gap of 50 bp  
255939 260942: contig of 5004 bp in length  
260943 261042: gap of unknown length  
261043 262102: contig of 1060 bp in length  
262103 262202: gap of unknown length  
262203 263692: contig of 1490 bp in length  
263693 263792: gap of unknown length  
263793 264824: contig of 1032 bp in length  
264825 264924: gap of unknown length  
264925 266484: contig of 1560 bp in length  
266485 266584: gap of unknown length  
266585 267818: contig of 1234 bp in length  
267819 267918: gap of unknown length  
267919 269286: contig of 1368 bp in length  
269287 269386: gap of unknown length  
269387 270427: contig of 1041 bp in length

Query Match 80.8%; Score 20.2; DB 14; Length 290052;  
Best Local Similarity 88.0%; Pred. No. 92;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
|||||

DB 199963 AAAACACACCACTGAGCTGTG 199987

RESULT 24  
CR854901/c  
LOCUS  
DEFINITION

CR854901 192168 bp DNA linear HTG 19-MAY-2005  
Danio rerio clone DKBY-5M2, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 11  
unordered pieces.

```
ACCESSION CR854901
VERSION CR854901.3 GI:66351706
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          1 (bases 1 to 192168)

REFERENCE
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
          On May 20, 2005 this sequence version replaced gi:54695370.

COMMENT ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfsh-help@sanger.ac.uk
          ----- Project Information
          Center project name: zKSM2
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 190046 bases at least Q40
          Consensus quality: 190568 bases at least Q30
          Consensus quality: 190823 bases at least Q20
          Insert size: 191168; sum-of-contigs
          Insert size: 191878; 5.8% error; agarose-fp
          Quality coverage: 6.16x in Q20 bases; sum-of-contigs Quality
          coverage: 6.35x in Q20 bases; agarose-fp
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence.
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 1 2172: contig of 2172 bp in length
          * 2173 2272: gap of 100 bp
          * 2273 46258: contig of 43986 bp in length
          * 46259 46359: gap of 100 bp
          * 46359 52193: contig of 5835 bp in length
          * 52194 52293: gap of 100 bp
          * 52294 52347: contig of 2954 bp in length
          * 52348 55347: gap of 100 bp
          * 55348 62364: contig of 7017 bp in length
          * 62365 62465: gap of 100 bp
          * 62465 68493: contig of 6028 bp in length
          * 68493 68593: gap of 100 bp
          * 68593 88001: contig of 19409 bp in length
          * 88002 88101: gap of 100 bp
          * 88102 107897: contig of 19796 bp in length
          * 107898 107998: gap of 100 bp
          * 107998 135043: contig of 27046 bp in length
          * 135044 135144: gap of 100 bp
          * 135144 152062: contig of 16919 bp in length
          * 152063 152163: gap of 100 bp
          * 152163 192168: contig of 40006 bp in length.

FEATURES
source
  1..192168
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /db_xref="taxon:7955"
    /clone="PKY-5M2"
    /clone_lib="DanioKey"
  misc_feature
    1..2172
    /note="assembly fragment:00037
    fragment_chain:1"
  misc_feature
    2273..46258
    /note="assembly fragment:01455
    fragment_chain:1"
  misc_feature
    46359..52193
    /note="assembly fragment:00049
    fragment_chain:1"
  misc_feature
    52294..55247
    /note="assembly fragment:00027
    fragment_chain:2"
  misc_feature
    55348..62364
    /note="assembly fragment:00161
    fragment_chain:2"
  misc_feature
    62465..68492
    /note="assembly fragment:00099
    fragment_chain:2"
  misc_feature
    68593..88001
    /note="assembly fragment:00366
    fragment_chain:2"
  misc_feature
    88102..107897
    /note="assembly fragment:00555
    fragment_chain:2"
  misc_feature
    107998..135043
    /note="assembly fragment:00752
    fragment_chain:2"
  misc_feature
    135144..152062
    /note="assembly fragment:00233
    fragment_chain:2"
  misc_feature
    152163..192168
    /note="assembly fragment:01010
    fragment_chain:2
    clone_end:SP6
    vector_side:right"

ORIGIN
Query Match 79.2%; Score 19.8; DB 14; Length 192168;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACTGAGCTG 23
   |||||
DB 80772 AACAAACCACTGAGCTG 80750

RESULT 25
BX469934 225586 bp DNA linear VRT 25-SEP-2003
LOCUS Zebrafish DNA sequence from clone CH211-209P5, complete sequence.
DEFINITION BX469934
ACCESSION BX469934
VERSION BX469934.5 GI:35760789
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          1 (bases 1 to 225586)
REFERENCE
AUTHORS Wood,J.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
          On Sep 25, 2003 this sequence version replaced gi:35209448.

COMMENT ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfsh-help@sanger.ac.uk
          -----
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a VAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: En:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiyong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

[http://www.sanger.ac.uk/projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/projects/D_rerio/fishmask.shtml) CH211-209P5 is from a CHORI-211 BAC library

VECTOR: pTARBAC2.1.

#### FEATURES

source

```
1..225586
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-209P5"
/clone_lib="CHORI-211"
```

#### ORIGIN

```
Query Match      79.2%; Score 19.8; DB 5; Length 225586;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 AAAAACAACACCACTGAGCTG 23

Db 161991 AACAAACAACCACTGAGCTG 162013

#### RESULT 26

```
AC098538          255707 bp   DNA       linear   HTG 10-MAY-2003
LOCUS             Rattus norvegicus clone CH230-92N23, *** SEQUENCING IN PROGRESS
DEFINITION        ***, 11 unordered pieces.
```

AC098538 GI:30521657

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 255707)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyte,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowals,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milsavljivic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,K., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Puaao,M., Quintero,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Sosa,J., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished 2 (bases 1 to 255707)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 10, 2003 this sequence version replaced gi:25095373.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GHIX

Center clone name: CH230-92N23

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 213914 bases at least Q40

Consensus quality: 218379 bases at least Q30

Consensus quality: 221566 bases at least Q20

Estimated insert size: 232388; sum-of-contigs estimation



Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 79107: contig of 79107 bp in length  
\* 79108 79207: gap of unknown length  
\* 79208 85284: contig of 6077 bp in length  
\* 85284 85384: gap of unknown length  
\* 85385 187331: contig of 101947 bp in length  
\* 187332 187431: gap of unknown length  
\* 187432 234683: contig of 47252 bp in length  
\* 234684 234783: gap of unknown length  
\* 234784 235975: contig of 1192 bp in length  
\* 235976 236075: gap of unknown length  
\* 236076 237584: contig of 1509 bp in length  
\* 237585 237684: gap of unknown length  
\* 237685 239262: contig of 1578 bp in length  
\* 239263 239362: gap of unknown length  
\* 239363 241904: contig of 2542 bp in length  
\* 241905 242004: gap of unknown length  
\* 242005 247701: contig of 5697 bp in length  
\* 247702 247801: gap of unknown length  
\* 247802 251585: contig of 3784 bp in length  
\* 251586 251686: gap of unknown length  
\* 251687 255707: contig of 4022 bp in length.

FEATURES  
source

1. .255707  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-92N23"  
79108..79207  
/estimated\_length=unknown  
misc\_feature 81675..85284  
/note="wgs contig"  
85285..85384  
/estimated\_length=unknown  
gap 187332..187431  
/estimated\_length=unknown  
misc\_feature 187432..188989  
/note="wgs contig"  
234684..234783  
/estimated\_length=unknown  
gap 235976..236075  
/estimated\_length=unknown  
gap 237585..237684  
/estimated\_length=unknown  
gap 239263..239362  
/estimated\_length=unknown  
gap 241905..242004  
/estimated\_length=unknown  
gap 247702..247801  
/estimated\_length=unknown  
gap 251586..251686  
/estimated\_length=unknown

ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 255707;  
Best Local Similarity 91.3%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 AAAAAACACCACTGAGCTGG 24  
|||||  
Db 223861 AAAAAACATCACCACTGGCTGG 223883

RESULT 27  
CHKHPX

LOCUS CHKHYPX 1288 bp DNA linear VRT 19-JUL-1995  
DEFINITION Gallus gallus hypothetical exon, complete cds.  
ACCESSION M84354  
VERSION M84354.1 GI:211948  
KEYWORDS homeobox protein.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 1288)  
AUTHORS Goldberg, G.S. and Kaczmarczyk, W.  
TITLE Sequence of a novel chicken genomic DNA fragment that hybridizes to  
the murine Hox-3.1 homeobox  
JOURNAL Gene 121 (2), 397-398 (1992)  
PUBMED 1359990

COMMENT Original source text: Gallus gallus DNA.  
FEATURES  
Location/Qualifiers

1. .1288  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9031"  
21..25  
/note="putative"  
306..316  
/note="putative"  
/rpt\_type=tandem  
343..353  
/note="putative"  
/rpt\_type=tandem  
372..380  
/note="putative"  
384..758  
/note="hypothetical exon homologous to murine Hox-3.1  
homeobox; putative"  
/codon\_start=1  
/label=ORF  
/protein\_id="AAA70193.1"  
/db\_xref="GI:211949"  
/translation="MSACALPARVGLCKWASGPSRTAAAEYRPNRYVSLPRLDPATY  
DTPLKRAERAEIRAIRLKRQYLQLNTPKPRVIVSGAGGAGGLWASPPHTDGS  
GPRQSLSPSRGLAGGAVFLSL"  
384..754  
/note="putative"  
755..799  
/note="putative"  
800..965  
/note="putative"  
1279..1284  
/note="putative"

TATA\_signal

repeat\_unit

repeat\_unit

GC\_signal

CDS

exon

intron

exon

polyA\_signal

ORIGIN

Query Match 77.6%; Score 19.4; DB 5; Length 1288;  
Best Local Similarity 95.2%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGC 21  
|||||

Db 196 AAAAAACACCACTCTGAGC 216  
|||||

RESULT 28  
GGWALTER

LOCUS GGWALTER 1288 bp DNA linear VRT 17-FEB-1997  
DEFINITION G.gallus DNA for hypothetical protein walter.  
ACCESSION X60778  
VERSION X60778.1 GI:517249  
KEYWORDS Gallus gallus (chicken)  
SOURCE Gallus gallus  
ORGANISM Gallus gallus



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archostralia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 1288)  
Goldberg, G.S. and Kaczmarczyk, W.  
A chicken genomic DNA fragment that hybridizes to the murine Hox-3.1 homeobox is likely to encode the NADH ubiquinone oxidoreductase subunit B15  
Gene 133 (2), 233-235 (1993)  
7901127

2  
Goldberg, G.S.  
Direct Submission  
Submitted (20-MAR-1991) G.S. Goldberg, University of Hawaii at Manoa, Cancer Research Centre of Hawaii, 1236 Lauhala Street, Honolulu, Hawaii 96813, USA

3 (bases 1 to 1288)  
Goldberg, G.  
Direct Submission  
Submitted (24-JUL-1991) Goldberg G., Cancer Research Center of Hawaii, University of Hawaii at Manoa, 1236 Lauhala Street, Honolulu, Hawaii 96813, USA

REMARK  
FEATURES  
source  
1. .1288  
/location="Qualifiers  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9031"  
1  
/citation=[2]  
/replaces="aaattaacacgtg"  
293. 303  
/standard\_name="alternating P / Y sequence"  
/notes="5-prime GAGCGCGTGC tandem repeat"  
/citation=[2]  
330. 340  
/standard\_name="Alternating P / Y sequence"  
/note="potential  
3-prime GAGCGCGTGC repeat"  
/citation=[2]  
359. 368  
/notes="This GC box is present as an overlapping inverted repeat"  
/citation=[2]  
371  
/notes="potential"  
/citation=[2]  
join(384. 742, 788. 968)  
/notes="potential open reading frame"  
/codon\_start=1  
/product="hypothetical protein walter"  
/protein\_id="CAA43193.1"  
/db\_xref="GI:517250"  
/db\_xref="GOA:P48306"  
/translation="MSKALPASVGLCKWAGSPRTAAEYRPNRYVSLPAELDPATY DTPLEKRAERLAIARLKRQYLQLNTPKPRVIVSAGAGAGGLWASPHPTDGS GPRQSLSPSLAGGAVPTGSRWGAAGWICWASEPMLWASRFCTLRILPCSGT MPRGTTCTPLSARHPRPS"  
384. 742  
/notes="potential"  
/citation=[2]  
743. 787  
/notes="potential intron"  
/citation=[2]  
788. 968  
/notes="potential"  
/citation=[2]  
1271  
/notes="potential Poly-A signal"  
/citation=[2]

ORIGIN

Query Match 77.6%; Score 19.4; DB 5; Length 1288;  
Best Local Similarity 95.2%; Pred. No. 2.1e-02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGC 21  
DB 196 AAAAAACACCACTGAGC 216

RESULT 29  
BX548018/c  
LOCUS  
DEFINITION  
Zebrafish DNA sequence from clone DKXY-23018 in linkage group 14, complete sequence.  
ACCESSION  
BX548018  
VERSION  
BX548018.9 GI:57834165  
KEYWORDS  
HTG  
SOURCE  
Danio rerio (zebrafish)  
ORGANISM  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 113040)  
Phillimore, B.  
Direct Submission  
Submitted (14-JAN-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jan 14, 2005 this sequence version replaced gi:51174232.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.  
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D\_rerio/fishmask.shtml DKXY-23018 is from a Zebrafish BAC library  
VECTOR: pIndigoBAC-5.  
Location/Qualifiers  
1. .113040  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKXY-23018"  
/clone\_lib="DanioKey"

FEATURES  
source  
1. .113040  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKXY-23018"  
/clone\_lib="DanioKey"

ORIGIN



RESULT 32  
AC153907/c

AL953906 172246 bp DNA linear ROD 31-JAN-2003  
Mouse DNA sequence from clone RP23-109E21 on chromosome 2, complete  
sequence.

ACCESSION AL953906  
VERSION AL953906.6 GI:28193313  
KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 172246)  
Chapman, J.

TITLE Direct Submission  
JOURNAL Submitted (31-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Feb 1, 2003 this sequence version replaced GI:27902085.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
from the RPCI-23 Mouse BAC Library  
constructed by the group of Pieter de Jong.  
For further details see <http://www.choi.org/bacpac/home.htm>  
VECTOR: pBACE3.6

#### FEATURES

source Location/Qualifiers  
1..172246  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone="RP23-109E21"  
/clone\_lib="RPCI-23"

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 9; Length 172246;  
Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACACACCACTGAGC 21  
|||||  
Db 87236 AAAAAACACCACTGAGC 87256

RESULT 32  
AC153907/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

AC153907 196286 bp DNA linear ROD 18-DEC-2004  
Mus musculus BAC RP23-205P6 (Roswell Park Cancer Institute  
(C57BL/6J Female) Mouse BAC Library) complete sequence.

AC153907

AC153907.1 GI:56710431

HTG.

Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 196286)

Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P.,  
Aredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,  
Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,  
Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,  
Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,  
Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,  
Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De  
Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,  
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K., Fa, M.,  
Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,  
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,  
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,  
Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,  
Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haeberlen, K.,  
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,  
Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J.,  
Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B.,  
Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,  
Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S.,  
Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F.,  
Leal, S., Lee, K., Lee, S., Legall, F. I., Lemon, S., Lewis, B., Li, B.,  
Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D.,  
London, P., Lopez, J., Lorenshew, L., Lozano, R., Luk, T., Madu, R.,  
Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,  
McClelland, H., McPherson, J., Mercadao, C., Metzker, M.,  
Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,  
Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Bastuagh, E.,  
Nott, A., Nwaokwelu, O., Obregon, M., Ochi-Okorie, C., Odeh, E.,  
Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B.,  
Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,  
Primus, E., Pul, I., Puazo, M., Qin, X., Quinn, A., Quiroz, J.,  
Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,  
Rives, C., Rodriguez, P., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W.,  
Santibanez, J., Santos, R., Savary, G., Scherer, S., Shen, H., Shen, Y.,  
Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,  
Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R.,  
Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villaseana, D., Virk, D.,  
Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,  
Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,  
Wu, J., Yakub, S., Yan, K., Yaun, Y., Yu, F., Zhang, J., Zhang, L.,  
Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 196286)

Worley, K.C.

Direct Submission

Submitted (18-DEC-2004)

Human Genome Sequencing Center, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Sequencing is completed to a minimum standard of double strand  
coverage with a minimum of 2 clones and 2 reads with no ambiguities  
or 2 chemistries with a minimum of 2 clones and 3 reads with no  
ambiguities. If the sequence quality does not meet this standard,  
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui

Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

source

1. 196286

/organism="Mus musculus"

/mol\_type="Genomic DNA"

/db\_xref="taxon:10090"

/clone="RP23-205F6"

complement(1. .78743)

/note="overlaps bases 1. .78743 of clone AC130816"

/function="clone overlap"

1488. .2766

/rpt\_family="L1\_Mm"

complement(2955. .3036)

/rpt\_family="ID4"

complement(3049. .3183)

/rpt\_family="B1F1"

3926. .3948

/rpt\_family="(TTTG)n"

4097. .4144

/rpt\_family="(TG)n"

complement(4630. .4766)

/rpt\_family="MIR"

complement(6227. .6412)

/rpt\_family="Lx2"

complement(6421. .6574)

/rpt\_family="B3"

8671. .9251

/rpt\_family="Lx2"

complement(10652. .10760)

/rpt\_family="L2"

10941. .10971

/rpt\_family="AT\_rich"

complement(11193. .11308)

/rpt\_family="B1\_Mur1"

11675. .11946

/rpt\_family="B4A"

complement(12897. .13134)

/rpt\_family="LIME3A"

13835. .13970

/rpt\_family="B1\_Mur2"

complement(14282. .14431)

/rpt\_family="L1\_Mus2"

complement(15215. .15386)

/rpt\_family="MIRb"

complement(15663. .15962)

/rpt\_family="LIMB7"

15979. .16006

/rpt\_family="(TTTA)n"

complement(16068. .16208)

/rpt\_family="LIMB5\_3end"

16223. .16326

/rpt\_family="CT-rich"

complement(16351. .16954)

/rpt\_family="LIMB7"

16957. .17120

/rpt\_family="B3A"

17131. .17151

/rpt\_family="AT\_rich"

complement(17207. .17275)

/rpt\_family="ID4"

18154. .18186

/rpt\_family="(TGGG)n"

18311. .18495

/standard\_name="32.MMEAP4PRC11.seq"

18622. .18671

/rpt\_family="L2"

18807. .18986

/rpt\_family="B3A"

19030. .19095

repeat\_region  
complement(19096. .19306)  
/rpt\_family="Lx7"  
19401. .19503  
/rpt\_family="(TTCCC)n"  
19508. .19535  
/rpt\_family="AT\_rich"  
19541. .19581  
/rpt\_family="CT-rich"  
19586. .19910  
/rpt\_family="Lx7"  
19919. .20070  
/rpt\_family="B3A"  
20144. .20172  
/rpt\_family="AT\_rich"  
20575. .20749  
/rpt\_family="MLTIC"  
21236. .21371  
/rpt\_family="MLTIC"  
21927. .21978  
/rpt\_family="L1Md\_F2"  
21996. .22144  
/rpt\_family="L1Md\_F2"  
22204. .22287  
/rpt\_family="AT\_rich"  
22387. .22486  
/rpt\_family="L2"  
complement(22774. .22852)  
/rpt\_family="ID2"  
24586. .24874  
/rpt\_family="B4"  
25007. .25529  
/rpt\_family="L1\_Mus1"  
complement(26134. .26398)  
/rpt\_family="L2"  
complement(27104. .27271)  
/rpt\_family="MIR"  
complement(27761. .27898)  
/rpt\_family="B1\_Mur1"  
complement(28154. .28228)  
/rpt\_family="ID4"  
complement(28976. .29022)  
/rpt\_family="L2"  
31282. .31409  
/rpt\_family="MTD"  
31441. .31587  
/rpt\_family="LIMC3"  
complement(31896. .31966)  
/rpt\_family="ID4"  
32349. .32371  
/rpt\_family="(TTTTTC)n"  
complement(32377. .32523)  
/rpt\_family="B1\_Mus1"  
32906. .32952  
/rpt\_family="(TG)n"  
complement(33229. .33408)  
/rpt\_family="HAL1"

Query Match 77.6%; Score 19.4; DB 9; Length 196286;  
Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAACACCACCTGAGCT 22

DB 176325 ACAAACACCACTGAGCT 176305

RESULT 33

CR847836

LOCUS

DEFINITION

SEQUENCE, 9 unordered pieces.

ACCESSION

CR847836

205309 bp DNA linear HTG 10-JUN-2005

Danio rerio chromosome 14 clone DKEY-285u18, WORKING DRAFT

CR847836

```

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

CR847836.4 GI:67509019
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 205309)
McLaren S.
Direct Submission
Submitted (08-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D_rerio/faq.shtmldataeight
On Jun 10, 2005 this sequence version replaced gi:54019838.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zk285J18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 202113 bases at least Q40
Consensus quality: 202607 bases at least Q30
Consensus quality: 202822 bases at least Q20
Insert size: 204509; sum-of-contigs
Insert size: 200755; 2.8% error; agarose-fp
Quality coverage: 7.24x in Q20 bases; sum-of-contigs Quality
coverage: 7.37x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 22499: contig of 22499 bp in length
* 22500: gap of 100 bp
* 22600: contig of 12387 bp in length
* 34987: gap of 100 bp
* 35087: contig of 90273 bp in length
* 125360: gap of 100 bp
* 125360: contig of 11031 bp in length
* 136491: gap of 100 bp
* 136591: contig of 10187 bp in length
* 146778: gap of 100 bp
* 146878: contig of 18039 bp in length
* 164917: gap of 100 bp
* 165017: contig of 24892 bp in length
* 189909: gap of 100 bp
* 190009: contig of 12411 bp in length
* 202420: gap of 100 bp
* 202520: contig of 2790 bp in length.
-----
Location/Qualifiers
1. .205309
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="14"
/clone="DKEY-285J18"
/clone_lib="DanioKey"
1. .22499
/note="assembly fragment:00627
fragment_chain:1"
22600. .34986
/note="assembly fragment:02675
fragment_chain:1"
35087. .125359

misc_feature
/note="assembly_fragment:01305
fragment_chain:1"
125460. .136490
/note="assembly_fragment:00001
fragment_chain:1"
136591. .146777
/note="assembly_fragment:02544
fragment_chain:1"
146878. .164916
/note="assembly_fragment:00376
fragment_chain:2"
165017. .189908
/note="assembly_fragment:00911
fragment_chain:2"
190009. .202419
/note="assembly_fragment:00183.0"
202520. .205309
/note="assembly_fragment:02526"

misc_feature
Query Match 77.6%; Score 19.4; DB 14; Length 205309;
Best Local Similarity 95.2%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAACAACACCACTGAGCT 22
| | | | | | | | | | | | | | | | | |
Db 3982 ATAACAACACCACTGAGCT 4002

RESULT 34
CR545469 214525 bp DNA linear HTG 16-APR-2005
LOCUS
DEFINITION
Danio rerio clone DKEY-151K14, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
CR545469
VERSION
CR545469.5 GI:62719142
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
McLaren S.
Direct Submission
Submitted (15-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 17, 2005 this sequence version replaced gi:50299899.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zk151K14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 209417 bases at least Q40
Consensus quality: 210277 bases at least Q30
Consensus quality: 210876 bases at least Q20
Insert size: 213525; sum-of-contigs
Insert size: 197611; 1.1% error; agarose-fp
Quality coverage: 9.62x in Q20 bases; sum-of-contigs Quality
coverage: 10.82x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

\* as soon as it is available and the accession number will  
\* be preserved.

1 38536: contig of 38536 bp in length  
\* 38537 gap of 100 bp  
\* 38637 contig of 2457 bp in length  
\* 41093: contig of 100 bp  
\* 41094 gap of 100 bp  
\* 41194 contig of 4145 bp in length  
\* 45338: contig of 100 bp  
\* 45339 gap of 100 bp  
\* 45438: contig of 7242 bp in length  
\* 52680: gap of 100 bp  
\* 52681 contig of 100 bp  
\* 52781 contig of 67029 bp in length  
\* 119810: contig of 100 bp  
\* 119909: gap of 100 bp  
\* 130198: contig of 10289 bp in length  
\* 130298: gap of 100 bp  
\* 130299 contig of 24751 bp in length  
\* 155049: gap of 100 bp  
\* 155050 contig of 2336 bp in length  
\* 155150: contig of 100 bp  
\* 157486: contig of 100 bp  
\* 157586: contig of 19364 bp in length  
\* 176949: gap of 100 bp  
\* 176950 contig of 17785 bp in length  
\* 177050: contig of 100 bp  
\* 194835: gap of 100 bp  
\* 194935: contig of 19591 bp in length.

FEATURES

Source

1..214525  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-151K14"  
/clone\_lib="DanioKey"  
1..38536  
/note="assembly\_fragment:01440  
/fragment\_chain:1"  
38637..41093  
/note="assembly\_fragment:00014  
/fragment\_chain:1"  
41194..45338  
/note="assembly\_fragment:00046  
/fragment\_chain:1"  
45439..52680  
/note="assembly\_fragment:00080  
/fragment\_chain:2"  
52781..119809  
/note="assembly\_fragment:02130  
/fragment\_chain:2"  
119910..130198  
/note="assembly\_fragment:00115  
/fragment\_chain:3"  
130299..155049  
/note="assembly\_fragment:00540  
/fragment\_chain:3"  
155150..157485  
/note="assembly\_fragment:00001"  
157586..176949  
/note="assembly\_fragment:00245.0"  
177050..194834  
/note="assembly\_fragment:00878.0"  
194935..214525  
/note="assembly\_fragment:00878.1"

ORIGIN

Query Match 77.6%; Score 19.4; DB 14; Length 214525;  
Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AACACCCACCTGAGCTGG 25

Db 162320 AACACCCACCTGAGCTGTG 162340

RESULT 35

AC105669

LOCUS

AC105669

DNA

261112 bp

linear

HTG 08-OCT-2002

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Rattus norvegicus clone CH230-64A8, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
AC105669  
AC105669.3 GI:22857190  
HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.  
1 (bases 1 to 261112)  
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anylebechechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorensuhewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaokelemeh,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
Pudzo,W., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanli,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

On Sep 14, 2002 this sequence version replaced gi:21736569.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GNMF  
Center clone name: CH230-64A8  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 229400 bases at least Q40  
Consensus quality: 232950 bases at least Q30  
Consensus quality: 235263 bases at least Q20  
Estimated insert size: 253424; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 261112: contig of 261112 bp in length.

## FEATURES

source  
misc\_feature 1..1063  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-64A8"  
misc\_feature 3350..5003  
/note="wgs contig"  
misc\_feature 136644..137706  
/note="wgs contig"

## ORIGIN

Query Match 77.6%; Score 19.4; DB 14; Length 261112;  
Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAACAACACCACTGAGCT 22

Db 38301 AAAAAAACCACCACTGAGCT 38321

RESULT 36

AF302834 AF302834 836 bp DNA linear INV 17-DEC-2000  
LOCUS Haliotis rubra clone 2.29 microsatellite VNTR sequence.  
DEFINITION Haliotis rubra clone 2.29 microsatellite VNTR sequence.  
ACCESSION AF302834  
VERSION AF302834.1 GI:11875757  
KEYWORDS  
SOURCE Haliotis rubra

## ORGANISM

Haliotis rubra  
Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;  
Veligastropoda; Haliotidae; Haliotidae; Haliotis.

## REFERENCE

1 (bases 1 to 836)

## AUTHORS

Evans,B.S.

## TITLE

Microsatellite marker conservation in the genus Haliotis

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 836)

## AUTHORS

Evans,B.S.

## TITLE

Direct Submission

## JOURNAL

Submitted (06-SEP-2000) Aquaculture and Biotechnology, CSIRO Marine  
Research, GPO Box 1538, Hobart, Tasmania 7001, Australia

## FEATURES

Location/Qualifiers

## source

1..836  
/organism="Haliotis rubra"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:36100"  
/clone="2.29"  
218..238  
322..437  
/note="microsatellite VNTR"  
/rpt\_type=tandem  
complement(519..538)

## primer\_bind

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 836;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACAACCACTGAGCTG 24

Db 633 AAAAAACAACCACTGAGCTTG 656

## RESULT 37

## LOCUS

## DEFINITION

Human DNA sequence from clone RP11-100N5 on chromosome 9 Contains  
part of the AKAP2 (PALM2) gene for A kinase (PRKA) anchor protein 2  
(paralemmn 2), complete sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AL627225 3994 bp DNA linear PRI 18-MAY-2005  
AL627225.11 GI:17066061  
HTG; AKAP2; anchor protein; CpG island; PALM2; paralemmn.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 3994)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Nov 25, 2001 this sequence version replaced gi:17017834.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
in the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr9>  
RP11-100N5 is from the library RPI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)





CDS	TNAQENSLADFLPQTPTDNPSEGRGEGVSKSFSDHGFYSPSSTLGDSPLVDDPLE														
	YQAGLLVQNAIQAIQAEQVDKAVSKTSRDGAEOQPEATVEEAAAFGEKQPSMF														
	PPQVSSPVQEKRDVLKILPAEDRALRERGPPQPLPAVOPSPINMEETRPBGVSFSK														
	YSAEALRSTASLLATQESDVMVGPFLKRSRKORTLSMEEIRAAQEREELKQRQ														
	VLQSTQSPRTKNAPSLRSTCYKATPGKLEKVPKPPPTTGGPSLPQDFDAPAEAAQTQ														
	RPKNLQTMLETYETHSKRRERMDSSYTKLLSCKVTSEVLEATRVNRRKSALALR														
	WEAGIYANOEEEDNE"														
	Join (AL158829.14:38598..38642,AL158829.14:125600..125680,														
	3937..3994,AL353806.9:1999..2071,AL353806.9:45125..45239,														
	AL353806.9:46451..46472,AL353806.9:53317..53418,														
Query Match	/gene="AKAP2"														
	/locus_tag="RP11-470J20.3-001"														
	/standard_name="OTTHUMP0000021890"														
	/note="match: proteins: Q96DUI"														
	/codon_start=1														
	/product="A kinase (PRKA) anchor protein 2"														
	76.8%;	Score 19.2;	DB 8;	Length 3994;											
	Best Local Similarity	87.5%;	Pred. No. 2.7e+02;												
	Matches 21;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;										
Db	2 AAAAAACAACCACTGAGCTGG 25														
	2439 AAAAAAACAACCACTGCGTGG 2416														
RESULT 38	AC104180														
	AC104180/c														
	LOCUS														
	DEFINITION														
	Leishmania major strain 90100 bp DNA linear HTG 26-MAR-2003														
	SEQUENCING IN PROGRESS ***, 2 ordered pieces.														
	AC104180														
	AC104180.5 GI:29244611														
	HTG; HTGS_PHASE2.														
	Leishmania major														
REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;														
	Leishmania.														
	1 (bases 1 to 90100)														
	Mylers,P.J., Sisk,E., Nelson,S., Mack,J., Vogt,C., Robertson,L.,														
	Worthey,E., Ivens,A., Marty,A., Munden,H., Stuart,K., Seyler,A. and														
	Rinta,J.														
	Direct Submission														
	Submitted (06-DEC-2001) Seattle Biomedical Research Institution, 4														
	Nickerson Street, Seattle, WA 98109-1651, USA														
	2 (bases 1 to 90100)														
TITLE	Mylers,P.J., Sisk,E., Nelson,S., Mack,J., Vogt,C., Nelson,S.,														
	Worthey,E., Ivens,A., Marty,A., Munden,H., Stuart,K., Seyler,A. and														
	Rinta,J.														
	Direct Submission														
	Submitted (26-MAR-2003) Seattle Biomedical Research Institution, 4														
	Nickerson Street, Seattle, WA 98109-1651, USA														
	On Mar 26, 2003 this sequence version replaced gi:25447632.														
	* NOTE: This is a 'working draft' sequence. It currently														
	* consists of 2 contigs. Gaps between the contigs														
	* are represented as runs of N. The order of the pieces														
COMMENT	* is believed to be correct as given, however the sizes														
	* of the gaps between them are based on estimates that have														
	* provided by the submitter.														
	* This sequence will be replaced														
	* by the finished sequence as soon as it is available and														
	* the accession number will be preserved.														
	* 1 68517: contig of 68517 bp in length														
	* 68518 68650: gap of unknown length														
	* 68651 90100: contig of 21450 bp in length.														
	Location/Qualifiers														
FEATURES	1..90100														
	/organism="Leishmania major"														
	/mol_type="genomic DNA"														
	/strain="Friedlin"														
	/db_xref="taxon:5664"														
	/chromosome="22"														
Gap	/clone="LB00273"														
	68518..68650														
	/estimated_length=unknown														
	ORIGIN	Query Match													
Best Local Similarity															
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;															
QY		1 AAAAAACAACCACTGAGCTGG 24													
	15628 AAAAAAGAACCACCGGCGG 15605														
	Db	RESULT 39													
AL359978.0/c															
WPCOMMENT															
Sequence split into 4 fragments LOCUS AL359978 Accession AL359978															
Fragment Name Begin End															
AL359978.0 1 110000															
AL359978.1 100001 210000															
AL359978.2 200001 310000															
AL359978.3 300001 387346															
LOCUS AL359978 387346 bp DNA linear HTG 16-JUL-2001															
DEFINITION Homo sapiens chromosome 9 clone RP11-144G15, 54 unordered pieces.															
ACCESSION AL359978															
VERSION AL359978.16 GI:14575237															
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.															
SOURCE Homo sapiens (human)															
ORGANISM Homo sapiens															
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;															
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;															
Hominidae; Homo.															
1															
Plumb,B.															
Direct Submission															
Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,															
CB10 18A, UK. E-mail enquiries: humquery@sanger.ac.uk															
requests: clonerequest@sanger.ac.uk															
On Jun 28, 2001 this sequence version replaced gi:12331081.															
----- Genome Center															
Center: Sanger Centre															
Center code: SC															
Web site: http://www.sanger.ac.uk															
Contact: humquery@sanger.ac.uk															
----- Project Information															
Center project name: bal14G15															
----- Summary Statistics															
Assembly program: XGAP4; version 4.5															
Sequencing vector: plasmid, L08752; 100% of reads															
Chemistry: Dye-terminator Big Dye; 100% of reads															
Consensus quality: 364536 bases at least Q40															
Consensus quality: 372753 bases at least Q30															
Consensus quality: 377828 bases at least Q20															
Insert size: 382046; sum-of-contigs															
Insert size: 172123; 13.2% error; agarose-fp															
Quality coverage: 8.09x in Q20 bases; sum-of-contigs Quality															
coverage: 22.67x in Q20 bases; agarose-fp															
-----															
* NOTE: This is a 'working draft' sequence. It currently															
* consists of 54 contigs. The true order of the pieces															
* is not known and their order in this sequence record is															
* arbitrary. Gaps between the contigs are represented as															
* runs of N, but the exact sizes of the gaps are unknown.															
* This record will be updated with the finished sequence															
* as soon as it is available and the accession number will															
* be preserved.															
1 35539: contig of 35539 bp in length															
35639: gap of 100 bp															
35640 46944: contig of 11305 bp in length															
46945 47044: gap of 100 bp															
47045 66901: contig of 19857 bp in length															
67001: gap of 100 bp															
66902															

\* 67002 70175: contig of 3174 bp in length  
\* 70176 70275: gap of 100 bp  
\* 70276 104443: contig of 34168 bp in length  
\* 104444 104543: gap of 100 bp  
\* 104544 181879: contig of 77336 bp in length  
\* 181880 181979: gap of 100 bp  
\* 181980 195035: contig of 13056 bp in length  
\* 195036 195135: gap of 100 bp  
\* 195136 197887: contig of 2752 bp in length  
\* 197888 197987: gap of 100 bp  
\* 197988 200370: contig of 2383 bp in length  
\* 200371 200470: gap of 100 bp  
\* 200471 204730: contig of 4260 bp in length  
\* 204731 204830: gap of 100 bp  
\* 204831 211277: contig of 6447 bp in length  
\* 211278 211377: gap of 100 bp  
\* 211378 215636: contig of 4259 bp in length  
\* 215637 215736: gap of 100 bp  
\* 215737 219304: contig of 3568 bp in length  
\* 219305 219404: gap of 100 bp  
\* 219405 221607: contig of 2203 bp in length  
\* 221608 221707: gap of 100 bp  
\* 221708 226929: contig of 5222 bp in length  
\* 226930 227029: gap of 100 bp  
\* 227030 231643: contig of 4614 bp in length  
\* 231644 231743: gap of 100 bp  
\* 231744 235404: contig of 3661 bp in length  
\* 235405 235504: gap of 100 bp  
\* 235505 246225: contig of 10721 bp in length  
\* 246226 246325: gap of 100 bp  
\* 246326 253367: contig of 7042 bp in length  
\* 253368 253467: gap of 100 bp  
\* 253468 258279: contig of 4812 bp in length  
\* 258280 258379: gap of 100 bp  
\* 258380 261753: contig of 3374 bp in length  
\* 261754 261853: gap of 100 bp  
\* 261854 264531: contig of 2678 bp in length  
\* 264532 264631: gap of 100 bp  
\* 264632 270234: contig of 5603 bp in length  
\* 270235 270334: gap of 100 bp  
\* 270335 273988: contig of 3634 bp in length  
\* 273989 274088: gap of 100 bp  
\* 274089 277669: contig of 3681 bp in length  
\* 277670 277869: gap of 100 bp  
\* 277870 283666: contig of 5797 bp in length  
\* 283667 283766: gap of 100 bp  
\* 283767 286170: contig of 2404 bp in length  
\* 286171 286270: gap of 100 bp  
\* 286271 288379: contig of 2109 bp in length  
\* 288380 288479: gap of 100 bp  
\* 288480 291608: contig of 3129 bp in length  
\* 291609 291708: gap of 100 bp  
\* 291709 296427: contig of 4719 bp in length  
\* 296428 296527: gap of 100 bp  
\* 296528 302497: contig of 5970 bp in length  
\* 302498 302597: gap of 100 bp  
\* 302598 319723: contig of 17126 bp in length  
\* 319724 319823: gap of 100 bp  
\* 319824 322858: contig of 3035 bp in length  
\* 322859 322958: gap of 100 bp  
\* 322959 325608: contig of 2650 bp in length  
\* 325609 325708: gap of 100 bp  
\* 325709 329450: contig of 3742 bp in length  
\* 329451 329550: gap of 100 bp  
\* 329551 331984: contig of 2434 bp in length  
\* 331985 332085: gap of 100 bp  
\* 332086 334354: contig of 2270 bp in length  
\* 334355 334454: gap of 100 bp  
\* 334455 337261: contig of 2807 bp in length  
\* 337262 337361: gap of 100 bp  
\* 337362 339476: contig of 2115 bp in length  
\* 339477 339576: gap of 100 bp  
\* 339577 342596: contig of 3020 bp in length

\* 342597 342696: gap of 100 bp  
\* 342697 345224: contig of 2828 bp in length  
\* 345225 345624: gap of 100 bp  
\* 345625 347670: contig of 2046 bp in length  
\* 347671 347770: gap of 100 bp  
\* 347771 352927: contig of 5157 bp in length  
\* 352928 353027: gap of 100 bp  
\* 353028 355067: contig of 2040 bp in length  
\* 355068 355167: gap of 100 bp  
\* 355168 357536: contig of 2369 bp in length  
\* 357537 357636: gap of 100 bp  
\* 357637 365237: contig of 7601 bp in length  
\* 365238 365337: gap of 100 bp  
\* 365338 367721: contig of 2384 bp in length  
\* 367722 367821: gap of 100 bp  
\* 367822 370537: contig of 2716 bp in length  
\* 370538 370637: gap of 100 bp  
\* 370638 374038: contig of 3401 bp in length  
\* 374039 374138: gap of 100 bp  
\* 374139 376389: contig of 2251 bp in length  
\* 376390 376489: gap of 100 bp  
\* 376490 379106: contig of 2617 bp in length  
\* 379107 379206: gap of 100 bp  
\* 379207 382405: contig of 3199 bp in length  
\* 382406 382505: gap of 100 bp  
\* 382506 384798: contig of 2293 bp in length  
\* 384799 384898: gap of 100 bp  
\* 384899 387346: contig of 2448 bp in length.

FEATURES

source

1..387346  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-144G15"  
/clone\_lib="RPC1-11.1"  
1..35539  
/note="assembly fragment:05228  
fragment chain:1"  
35640..46944  
/note="assembly fragment:09756  
fragment chain:1"  
47045..66901  
/note="assembly fragment:00410  
fragment chain:1"  
67002..70175  
/note="assembly fragment:01425  
fragment chain:1"  
70276..104443  
/note="assembly fragment:09780  
fragment chain:1"  
104544..181879  
/note="assembly fragment:05688  
fragment chain:1"  
181980..195035  
/note="assembly fragment:03194  
fragment chain:1"  
195136..197887  
/note="assembly fragment:02620  
fragment chain:2"  
197988..200370  
/note="assembly fragment:03650  
fragment chain:2"  
200471..204730  
/note="assembly fragment:00487  
fragment chain:2"  
204831..211277  
/note="assembly fragment:04780  
fragment chain:2"  
211378..215636

Query Match 76.8%; Score 19.2; DB 14; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 2.8e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAACACCACTGAGCTGG 25  
||||| ||||| ||||| ||||| |||||

Db 104982 AAAAAACCACTGGCTGG 104959  
||||| ||||| ||||| ||||| |||||

RESULT 40  
AL359978 1/c  
WPCOMMENT

Sequence split into 4 fragments LOCUS AL359978 Accession AL359978

Fragment Name	Begin	End
AL359978_0	1	110000
AL359978_1	100001	210000
AL359978_2	200001	310000
AL359978_3	300001	387346

Continuation (2 of 4) of AL359978 from base 100001 (AL359978 Homo sapiens chromosome 9 c

Query Match 76.8%; Score 19.2; DB 14; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAACCACTGAGCTGG 25  
||||| ||||| ||||| ||||| |||||

Db 4982 AAAAAACCACTGGCTGG 4959  
||||| ||||| ||||| ||||| |||||

RESULT 41  
BX936368 2/c  
WPCOMMENT

Sequence split into 5 fragments LOCUS BX936368 Accession BX936368

Fragment Name	Begin	End
BX936368_0	1	110000
BX936368_1	100001	210000
BX936368_2	200001	310000
BX936368_3	300001	410000
BX936368_4	400001	429582

Continuation (3 of 5) of BX936368 from base 200001 (BX936368 Danio rerio clone DKEY-1891

Query Match 76.8%; Score 19.2; DB 14; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAACCACTGAGCTGG 25  
||||| ||||| ||||| ||||| |||||

Db 71918 AAAAAACCACTGCTGG 71895  
||||| ||||| ||||| ||||| |||||

RESULT 42  
CT005261\_0  
WPCOMMENT

Sequence split into 8 fragments LOCUS CT005261 Accession CT005261

Fragment Name	Begin	End
CT005261_0	1	110000
CT005261_1	100001	210000
CT005261_2	200001	310000
CT005261_3	300001	410000
CT005261_4	400001	510000
CT005261_5	500001	610000
CT005261_6	600001	710000
CT005261_7	700001	716508

LOCUS CT005261 716508 bp DNA linear HTG 04-JUL-2005

DEFINITION Leishmania major strain Friedlin, \*\*\* SEQUENCING IN PROGRESS \*\*\*.

ACCESSION CT005261 AC008240

VERSION CT005261.1 GI:68224038

KEYWORDS HTG; HTGS PHASE2.

SOURCE Leishmania major

ORGANISM Leishmania major

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

1 (bases 1 to 716508)

Peacock,C.S., Murphy,L., Ivens,A.C., Berriman,M., Blackwell,J.,

Smith,D., Collins,M., Foster,N., Harris,D., Oliver,K., O'Neil,S.,

Saunders,D., Seeger,K., Warren,T., Rajandream,M. and Barrall,B.G.  
Direct Submission  
Submitted (15-JUN-2005) Peacock C.S., The Wellcome Trust Sanger  
Institute Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
On Jul 6 2005 this sequence version replaced gi:29825871.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 716608: contig of 716608 bp in length.

FEATURES

Location/Qualifiers

1..716608

/organism="Leishmania major"

/mol\_type="genomic DNA"

/strain="Friedlin"

/db\_xref="taxon:5664"

/note="LmjF22\_01.20050601\_V5.2"

complement(3313..5241)

/locus\_tag="LmjF22.0010"

complement(3313..5241)

/locus\_tag="LmjF22.0010"

/note="LmjF22.0010. Predicted protein, len = 643 aa,  
unknown protein; predicted pi = 8.0047"

/codon\_start=1

/product="hypothetical protein, conserved"

/protein\_id="CAJ03760.1"

/db\_xref="GI:68224039"

/translation="MRGVDKGPGREGVEVDVAGVVTEDLTKDGVNDPNSLSQN  
EPGIEESDRYKRDATVHLKRTIIRNSVFLVLVGFNNPHYCLIMSAASLAEGY  
GLKLVALTWANIIFGIARVLNMEIATVSVNVPITAMALLSGLGIFPVSPAAPIG  
GYNVSAFVVMGLIGVLLGTASTYGGESVMLTFVQRPDSIVGAWSGTGSGVAASLI  
FGLTSAGTQQOTLVITPLCIITWLCFMLGNVSPRKVLATRTNGRDIIILCE  
GVEDAHICLLHERALHDVSVSEDTITKVMNIRAVSHQSPVHVHPSPSABEGDG  
LVYPCRCVCGCCSPNGRWRWDVGPDLIFMHTMAMWFFNLAVYAYAAQLMA  
PFSFYCEPEWKNFVKNVSVVVCQVQLGVLISRSLLIKVLPYGVVISIIQVNAV  
CHIVQARVMYIKNSKRVQGLSFILFAMWIFGLMGSLVNVVLILKRSITLRR  
EREAVEAYLTSKRDVSGRYAEDTELKKAADERAALVAKLEKPVSEIMPKPSQ  
SIASAAPQSQILGLATQERDAIKAIHCNLDNVWAAREMGMINIGALYATVGTITLG  
TVVDLIPTNTMLKHSKSC"

complement(join(3349..3408,3751..3819,3862..3930,  
3949..4008,4051..4119,4480..4548,4576..4644,4663..4731,  
4759..4827,4864..4923,4951..5019))

/locus\_tag="LmjF22.0010"

/notes="11 probable transmembrane helices predicted for  
LmjF22.0010 by TMHMM2.0 at aa 75-97, 107-126, 139-161,  
171-193, 200-222, 232-254, 375-397, 412-431, 438-460,  
475-497 and 612-631"

complement(3676..5025)

/locus\_tag="LmjF22.0010"

/notes="HMPfam

HMPfam hit to PF02487, CLN3 protein, score 1.7e-08"

complement(11185..11619)

/locus\_tag="LmjF22.0020"

complement(11185..11619)

/locus\_tag="LmjF22.0020"

/notes="LmjF22.0020, predicted protein, len = 643 aa,  
unknown; predicted pi = 5.5753"

/codon\_start=1

/product="hypothetical protein, unknown function"

/protein\_id="CAJ03762.1"

/db\_xref="GI:68224040"

/translation="MSSTVPHAAAKVVDCKEIGVITDLHSPPTASNFLAYNRVCSA  
LDSYLMQPIINEAQYDCAQVDSLNGPTSFYDHYVRKPTRKRLGAVGCESPMEHV  
GVSEDFLMPAQFDLPAAQACAFSDPLTRLATGRWPPRD"

complement(12239..12805)

/locus\_tag="LmjF22.0030"

complement(12239..12805)

/locus\_tag="LmjF22.0030"

misc\_feature

misc\_feature

gene

CDS

gene

CDS







```
/db_xref="UniProt/TREMBL:Q5T719"
/translation="WAEAEHKKERLQAIKAEKKRQTEIEGKQQLDEQILLQHSKSK
VLRKWLLOGIPAGTAEERARRQSEDEFVKQLEDNIOQLRLEQITLSEESQIS
AKEQILILEKLETESFKDFQGFSTDDGVDVNISSQLPDLPIICRSATSPSQDG
TSRAAVAMEINVEKDKTGTETKILSTIGPEGVHQGVKGVYDGGTKVYEVRSGG
TVVNGVHKLSTKVEELIQAGQSGLGCHVSERTIADGSLSPKHEHMLCKEAKLE
MYHKSRKHDSGNPQQOQAAPAEANLDQVTMIFMGYONIIDESETKKVLGYDE
TIKAEVLIDENDEKSKETVYDUSTIDGNAELVSGRPVSDTTEPSSPEGKESLA
TEPAFCVGNVILLKEGSGASNATETSGPDMTIKKPQLSE"
gene
join(AL158829.14:38418..38642,AL158829.14:125600..125680,
AL627225.11:13937..3994,1999..2071,45125..45239,
46451..46472,53317..53418,AL158823.11:28293..28378,
AL158823.11:148466..150878,AL158823.11:168658..168836,
AL158823.11:174433..174471,AL353598.12:7180..11293)
/genes="AKAP2"
/locus_tag="RP11-470J20.3-001"
mRNA
join(AL158829.14:38418..38642,AL158829.14:125600..125680,
AL627225.11:13937..3994,1999..2071,45125..45239,
46451..46472,53317..53418,AL158823.11:28293..28378,
AL158823.11:148466..150878,AL158823.11:168658..168836,
AL158823.11:174433..174471,AL353598.12:7180..11293)
/genes="AKAP2"
/locus_tag="RP11-470J20.3-001"
/product="A kinase (PKA) anchor protein 2"
/notes="match: ESTs: AA769109.1 AI753746.1 AW363024.1
AK489512.1 BB677276.1 BG171898.1 BG346015.1 BM970671.1
BQ212373.1 BU070781.1 BU156689.1 BU682644.1
match: CDNAS: AB023137.1 AJ303079.1"
CDS
join(AL158829.14:38598..38642,AL158829.14:125600..125680,
AL627225.11:13937..3994,1999..2071,45125..45239,
46451..46472,53317..53418,AL158823.11:28293..28378,
AL158823.11:148466..150878,AL158823.11:168658..168836,
AL158823.11:174433..174471,AL353598.12:7180..7278)
/genes="AKAP2"
/locus_tag="RP11-470J20.3-001"
/standard_names="OTTHUMP0000021887"
/notes="match: proteins: CAC38839 O54931 Q99L95 Q9Y2D5"
/codon_start=1
/product="A kinase (PKA) anchor protein 2"
/protein_id="CAI12434.1"
/db_xref="GI:55958365"
/translation="WAEAEHKKERLQAIKAEKKRQTEIEGKQQLDEQILLQHSKSK
VLRKWLLOGIPAGTAEERARRQSEDEFVKQLEDNIOQLRLEQITLSEESQIS
AKEQILILEKLETESFKDFQGFSTDDGVDVNISSQLPDLPIICRSATSPSQDG
TSRAAGVGNVILLKEGSGASNATETSGPDMTIKKPQLSEDDIWLKSGDNYSAIIL
EPAAASLSPDKMEIEVSAECKVPGITSTPHMDHPSAFYSPPHNGLLTDHESL
DNDVAREIRYLDLEANCCDSADVDTYNGTSPSPGAVLVGLSPVPVHEATQPEPT
ERTASRAQAPPHIELSNSPDPMAEARTNGHSPQPRDALGDSLQVPVSPSTSSRC
```

```
Query Match 76.8%; Score 19.2; DB 8; Length 111051;
Best Local Similarity 87.5%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACACACCTGAGCTGG 25
||||| ||||| ||||| ||||| |||||
Db 445 AAAATAACACACCTGGCTGG 422

RESULT 45
AC129011 124350 bp DNA linear HTG 26-NOV-2002
LOCUS
DEFINITION
Leishmania major strain Friedlin chromosome 22 clone LB02569, ***
SEQUENCING IN PROGRESS ***, 3 ordered pieces.
AC129011
AC129011.4 GI:25447647
HTG: HTGS PHASE2.
Leishmania major
Leishmania major
Leishmania major
Leishmania.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 1 to 124350)
Myler,P.J., Sisk,E., Vogt,C., Robertson,L., Worthey,L., Munden,H.,
Nelson,S., Ivens,A., Seyler,A., Rinta,J. and Stuart,K.
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
Myler,P.J., Sisk,E., Vogt,C., Robertson,L., Worthey,L., Munden,H.,
Nelson,S., Ivens,A., Seyler,A., Rinta,J. and Stuart,K.
TITLE
JOURNAL
COMMENT
On Nov 26, 2002 this sequence version replaced gi:2218467.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 19919: contig of 19919 bp in length
* 19920 20074: gap of unknown length
* 20075 59697: contig of 39623 bp in length
* 59698 59851: gap of unknown length
* 59852 124350: contig of 64499 bp in length.
FEATURES
Location/Qualifiers
1..124350
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="22"
/clone="LB02569"
19920..20074
/estimated_length=unknown
59698..59851
/estimated_length=unknown

ORIGIN
Query Match 76.8%; Score 19.2; DB 14; Length 124350;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACACACCTGAGCTGG 24
||||| ||||| ||||| ||||| |||||
Db 2226 AAAAAGAACACACCGAGCGG 2249

RESULT 46
AC020922
LOCUS
DEFINITION
Homo sapiens chromosome 19 clone CTD-2105E13, complete sequence.
AC020922
AC020922.9 GI:27777576
HTG.
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 134793)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS
Direct Submission
TITLE
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 134793)
DOE Joint Genome Institute.
AUTHORS
Direct Submission
TITLE
Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
JOURNAL
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134793)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS
Direct Submission
TITLE
Submitted (01-AUG-2002) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL
Drive, Walnut Creek, CA 94598, USA
```





AC145018/c  
LOCUS AC145018 141663 bp DNA linear HTG 26-JUN-2003  
DEFINITION Felis catus clone RP86-279N10, WORKING DRAFT SEQUENCE, 5 ordered pieces.  
AC145018  
VERSION AC145018.2 GI:32261343  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE Felis catus (cat)  
ORGANISM Felis catus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae; Felinae; Felis.  
REFERENCE 1 (bases 1 to 141663)  
AUTHORS Antonellis A., Ayele K., Beckstrom-Sternberg, S.M., Benjamin, B., Blankesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K., Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masiello, C., Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K., Sison, C., Stantripop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V., Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.  
NISC Comparative Sequencing Initiative  
UNPUBLISHED  
2 (bases 1 to 141663)  
Green, E.D.  
Direct Submission  
Submitted (30-MAY-2003) NIH Intramural Sequencing Center, 8717  
JOURNAL Grovemont Circle, Gaithersburg, MD 20877, USA  
3 (bases 1 to 141663)  
Green, E.D.  
Direct Submission  
Submitted (26-JUN-2003) NIH Intramural Sequencing Center, 8717  
JOURNAL Grovemont Circle, Gaithersburg, MD 20877, USA  
COMMENT On Jun 26, 2003 this sequence version replaced gi:31193896.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: efl  
Center clone name: 279N10  
The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 140709 bases at least Q40  
Consensus quality: 140932 bases at least Q30  
Consensus quality: 141174 bases at least Q20  
Insert size: 123000; agarose-fp  
Insert size: 141263; sum-of-contigs  
Quality coverage: 11.81x in Q20 bases; agarose-fp  
Quality coverage: 10.28x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces \* is believed to be correct as given, however the sizes \*  
\* of the gaps between them are based on estimates that have \* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and \* the accession number will be preserved.  
\* 1 755: contig of 755 bp in length  
\* 756 855: gap of unknown length  
\* 856 8935: contig of 8080 bp in length  
\* 8936 9035: gap of unknown length  
\* 9036 32437: contig of 23401 bp in length  
\* 32437 32537: gap of unknown length  
\* 32537 55737: contig of 23201 bp in length  
\* 55738 55837: gap of unknown length  
\* 55838 141663: contig of 85826 bp in length.  
FEATURES  
source  
1..141663  
/organism="Felis catus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9685"  
/clone="RP86-279N10"  
/clone\_lib="RP86"  
1..755  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:left"  
756..855  
/estimated\_length=unknown  
856..8935  
/note="assembly\_fragment"  
8936..9035  
/estimated\_length=unknown  
9036..32436  
/note="assembly\_fragment"  
32437..32536  
/estimated\_length=unknown  
32537..55737  
/note="assembly\_fragment"  
55738..55837  
/estimated\_length=unknown  
55838..141663  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 141663;  
Best Local Similarity 87.5%; Pred. No. 2.9e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 3;  
QY 2 AAAAAACACCCACTGAGCTGGG 25  
|||||  
Db 124266 AAAAAACACCCACTGAGCAGGG 124243  
|||||  
RESULT 49  
AC157291  
LOCUS Bos taurus clone CH240-72B6, WORKING DRAFT SEQUENCE, 7 unordered pieces.  
DEFINITION AC157291 154763 bp DNA linear HTG 01-JUL-2005  
ACCESSION AC157291  
VERSION AC157291.2 GI:60301583  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 154763)  
AUTHORS Muzny, D., Marie, D., Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, P., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, W., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gragegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guerra, M., Guver, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, H., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulky, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokelemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 154763)  
Worley, K.C.  
Submitted (11-FEB-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 154763)  
Cow Genome Sequencing Consortium.  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 29, 2005 this sequence version replaced gi:59276096.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: FGCN  
Center clone name: CH240-7286  
Assembly program: Atlas 3.0;  
Consensus quality: 152641 bases at least Q40  
Consensus quality: 153258 bases at least Q30  
Consensus quality: 153807 bases at least Q20  
Estimated insert size: 156808; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2299: contig of 2299 bp in length  
\* 2300: gap of 83 bp  
\* 2382: contig of 9618 bp in length  
\* 12001: contig of 50 bp  
\* 12050: gap of 50 bp  
\* 12051: contig of 26562 bp in length  
\* 38613: contig of 153 bp  
\* 38765: gap of 153 bp  
\* 55717: contig of 16952 bp in length  
\* 55767: gap of 50 bp  
\* 99117: contig of 43350 bp in length  
\* 99168: gap of 50 bp  
\* 99168: gap of 50 bp  
\* 144531: contig of 45363 bp in length  
\* 144531: gap of 50 bp  
\* 144581: 154763: contig of 10183 bp in length.

FEATURES  
source

gap  
gap  
gap  
gap  
gap  
gap

ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 154763;  
Best Local Similarity 87.5%; Pred. No. 2.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAAACACCACTGAGCTGGG 25  
|||||  
Db 147405 AAAAAACACCACTGAGCTGGG 147428  
|||||  
RESULT 50  
AC016868/c 155691 bp DNA linear PRI 06-FEB-2002  
LOCUS Homo sapiens chromosome 8, clone RF11-15G16, complete sequence.  
DEFINITION AC016868  
ACCESSION AC016868  
VERSION AC016868.8 GI:18377299  
KEYWORDS HTG.

SOURCE	ORGANISM	REFERENCE	TITLE	REFERENCE	TITLE	FEATURES	
Homo sapiens (human)	Homo sapiens	1 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	3 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,L., Campopiano,A., Chang,J., Chazaro,B., Choepe,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glende,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research
Homo sapiens	Homo sapiens	2 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	3 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,L., Campopiano,A., Chang,J., Chazaro,B., Choepe,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glende,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	4 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	5 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	6 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	7 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	8 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tizrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Center for Genome Research		
Homo sapiens	Homo sapiens	9 (bases 1 to 155691)	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Caslie,A., Collangelo,M., Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McG				

```
repeat_region /rpt_family="L1ME4A"
17327..17569
/rpt_family="MER106A"
complement(17570..17793)
/rpt_family="L1ME4A"
complement(17794..18102)
/rpt_family="AluSx"
complement(18103..18407)
/rpt_family="L1ME4A"
complement(18409..18588)
/rpt_family="AluSg/x"
18589..18842
/rpt_family="AluSx"
complement(19365..19666)
/rpt_family="AluSx"
complement(19730..20148)
/rpt_family="L1ME3"
complement(20173..20384)
/rpt_family="MLT1D"
complement(20445..20505)
/rpt_family="MLT1D"
complement(20506..20804)
/rpt_family="AluSx"
21254..21280
/rpt_family="(CA)n"
21282..21314
/rpt_family="(TA)n"
complement(21320..21471)
/rpt_family="L1ME2"
complement(21472..21635)
/rpt_family="FRAM"
complement(21636..22021)
/rpt_family="L1ME2"
complement(22022..22063)
/rpt_family="THE1C"
22064..22374
/rpt_family="AluSx"
complement(22312..22316)
/note="<30 qual SNGL region"
complement(22337..22346)
/note="<30 qual SNGL region"
complement(22375..22746)
/rpt_family="THE1C"
complement(22747..22999)
/rpt_family="L1ME2"
complement(23000..23306)
/rpt_family="AluSq"
complement(23307..24065)
/rpt_family="L1ME2"
```

Query Match 76.8%; Score 19.2; DB 8; Length 155691;  
Best Local Similarity 87.5%; Pred.No. 2.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGAGCTGG 24  
||||| ||| |||||  
Db 25323 AAAAAAACCCCTCACTGAGCTGG 25300

Search completed: February 3, 2006, 21:26:00  
Job time : 877 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-3

Perfect score: 25

Sequence: 1 aaaaacaccaccactgagctggg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21.\*

1: geneseqn1980s.\*

2: geneseqn190s.\*

3: geneseqn2000s.\*

4: geneseqn2001bs.\*

5: geneseqn2002bs.\*

6: geneseqn2003bs.\*

7: geneseqn2004bs.\*

8: geneseqn2005bs.\*

9: geneseqn2006bs.\*

10: geneseqn2007bs.\*

11: geneseqn2008bs.\*

12: geneseqn2009bs.\*

13: geneseqn2010bs.\*

14: geneseqn2011bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	76.8	503	9	ACH44272 Human foe
C 2	18.8	75.2	42061	13	ABD33566 Human can
C 3	18.8	75.2	96898	10	ACH01384 Human BIV
C 4	18.6	74.4	90	8	ABZ09104 Human oil
C 5	18.6	74.4	90	10	ABZ78557 Tumour su
C 6	18.6	74.4	93	2	AAT24655 Human gen
C 7	18.6	74.4	382	10	ADD47465 Rat gene
C 8	18.6	74.4	540	4	ABA08904 Human oil
C 9	18.6	74.4	540	10	ADI21508 Novel hum
C 10	18.6	74.4	540	13	ADS98314 Protein E
C 11	18.6	74.4	591	4	AAH32100 Human oil
C 12	18.6	74.4	762	6	ABL39719 Human NS
C 13	18.6	74.4	777	6	ABZ42980 Human GPC
C 14	18.6	74.4	812	11	ACL13584 Rice abio
C 15	18.6	74.4	932	4	AAH32071 Human oil
C 16	18.6	74.4	950	8	ABX34731 Human mdd
C 17	18.6	74.4	951	10	ABZ77938 Human G p
C 18	18.6	74.4	1065	5	AAS15907 DNA encod
C 19	18.6	74.4	1069	6	ABK40204 Human G p

C 20	18.6	74.4	1232	13	ADS97926	Ad97926 Protein f
C 21	18.6	74.4	1329	6	ABN89112	Abn89112 Human GPC
C 22	18.6	74.4	1344	6	ABA95482	AbA95482 Human lam
C 23	18.6	74.4	1450	6	ABK92484	Abk92484 Human pro
C 24	18.6	74.4	2027	6	ABK95452	Abk95452 DNA encod
C 25	18.6	74.4	2028	6	ABK40203	Abk40203 Human G p
C 26	18.6	74.4	2312	8	ABZ58229	Abz58229 Human G-P
C 27	18.6	74.4	2464	4	AAK68477	Aak68477 Human imm
C 28	18.6	74.4	2464	4	AAK68479	Aak68479 Human imm
C 29	18.6	74.4	2464	4	AAK04040	Aal04040 Human rep
C 30	18.6	74.4	2464	4	AAK04041	Aal04041 Human rep
C 31	18.6	74.4	2775	12	ADI16291	Adi16291 Human nuc
C 32	18.6	74.4	3009	12	ADO64688	Ado64688 Novel hum
C 33	18.6	74.4	3029	12	ADQ24417	Adq24417 Human sof
C 34	18.6	74.4	3030	12	ADI16310	Adi16310 Human nuc
C 35	18.6	74.4	3154	13	ADR25447	Adr25447 Breast ca
C 36	18.6	74.4	3190	3	AAZ57865	Aaz57865 Protein r
C 37	18.6	74.4	3226	14	ADX05939	Adx05939 Cyclin-de
C 38	18.6	74.4	3512	13	ACN42607	Acn42607 Human dta
C 39	18.6	74.4	4281	13	ADR07844	Adr07844 Full leng
C 40	18.6	74.4	9840	4	AAK67429	Aak67429 Human imm
C 41	18.6	74.4	16578	10	ADI22003	Adi22003 Novel hum
C 42	18.6	74.4	16578	13	ADS98483	Ads98483 Protein f
C 43	18.6	74.4	39119	8	ADZ74034	Abz74034 Secreted
C 44	18.6	74.4	39119	8	ADA98641	Ada98641 Human sec
C 45	18.6	74.4	39119	10	ADC20764	Adc20764 Human sec
C 46	18.6	74.4	39119	10	ABZ67621	Abz67621 Human sec
C 47	18.6	74.4	53099	6	ABX98185	Abx98185 Human mul
C 48	18.6	74.4	86950	14	ADX98572	Adx98572 Human HTO
C 49	18.6	74.4	122923	11	ACN44026	Acn44026 Human gen
C 50	18.6	74.4	128993	8	ACF62749	Acf62749 Cancer ba
C 51	18.6	74.4	128993	8	ADB20868	Adb20868 MRP1 base
C 52	18.6	74.4	128993	10	ADB87957	Adb87957 Human UGT
C 53	18.6	74.4	128993	10	ADB96940	Adb96940 Human MDR
C 54	18.6	74.4	128993	10	ADB92131	Adb92131 Human MDR
C 55	18.6	74.4	177380	8	ACF62751	Acf62751 Cancer ba
C 56	18.6	74.4	177380	8	ADB20870	Adb20870 MRP1 base
C 57	18.6	74.4	177380	10	ADB87959	Adb87959 Human UGT
C 58	18.6	74.4	177380	10	ADB96942	Adb96942 Human MDR
C 59	18.6	74.4	177380	10	ADB92133	Adb92133 Human MDR
C 60	18.6	74.4	241748	14	ADZ13116	Adz13116 Murine ca
C 61	18.4	73.6	2097	8	ACA52966	ACA52966 Prokaryot
C 62	18.4	73.6	47756	12	ADQ59512	Adq59512 Human can
C 63	18.4	73.6	57036	14	ADZ13891	Adz13891 Human can
C 64	18.2	72.8	407	3	AAC22116	Aac22116 Human sec
C 65	18.2	72.8	598	4	ABA60937	AbA60937 Human foe
C 66	18.2	72.8	598	4	AAI40833	Aai40833 Probe #95
C 67	18.2	72.8	598	4	ABA28906	AbA28906 Probe #73
C 68	18.2	72.8	598	4	AAK35116	Aak35116 Human bon
C 69	18.2	72.8	598	4	AAK09227	Aak09227 Human bra
C 70	18.2	72.8	598	4	ABS34868	AbS34868 Human liv
C 71	18.2	72.8	598	6	ABS09572	AbS09572 Human gen
C 72	18.2	72.8	4624	4	ABL22501	AbL22501 Drosophil
C 73	18.2	72.8	110000	14	ADZ42274_2	Concentration (3 of
C 74	18.2	72.8	326014	6	ADQ94981	Adq94981 Human kin
C 75	18.2	72.8	326014	12	ABQ94981	AbQ94981 Human kin
C 76	17.8	71.2	108	4	ABA71547	AbA71547 Human foe
C 77	17.8	71.2	108	4	AAI51824	Aai51824 Probe #20
C 78	17.8	71.2	108	4	AAK45906	Aak45906 Human bon
C 79	17.8	71.2	108	4	AAK19876	Aak19876 Human bra
C 80	17.8	71.2	108	4	ABS45604	AbS45604 Human liv
C 81	17.8	71.2	108	6	ABS20191	AbS20191 Human gen
C 82	17.8	71.2	352	4	AAS26480	Aas26480 Human CDN
C 83	17.8	71.2	352	4	AAI62482	Aai62482 Human bre
C 84	17.8	71.2	352	4	AAK00448	Aal00448 Human rep
C 85	17.8	71.2	352	8	ABX73821	Abx73821 Human nov
C 86	17.8	71.2	360	4	ABA59020	AbA59020 Human foe
C 87	17.8	71.2	360	4	AAI38747	Aai38747 Probe #74
C 88	17.8	71.2	360	4	AAK32941	Aak32941 Human bon
C 89	17.8	71.2	360	4	AAK07190	Aak07190 Human bra
C 90	17.8	71.2	360	4	ABS32670	AbS32670 Human liv
C 91	17.8	71.2	360	6	ABS07747	AbS07747 Human gen
C 92	17.8	71.2	435	5	AAS67401	Aas67401 DNA encod





```
XX WO2004058146-A2.
XX
XX
XX 15-JUL-2004.
XX
XX 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-499109/47.
XX
XX Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.
XX
XX Claim 16; SEQ ID NO 770; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
XX associated (CA) nucleic acids encoding them. The invention also relates
XX to a method for treating cancers involving administering to a patient an
XX inhibitor of CAP, and a method of screening for anticancer activity in a
XX potential drug involving providing a cell that expresses a CA gene,
XX contacting a tissue sample derived from a cancer cell with an anticancer
XX drug candidate and monitoring the effect of the anticancer drug candidate
XX on expression of the CA gene. The CAP proteins are useful for detecting
XX cancer associated with expression of a CAP protein in a test cell sample
XX and for screening for a bioactive agent capable of modulating the
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX cancer, involving determining the expression of a CA nucleic acid in a
XX tissue. This sequence represents a human CA gene of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 42061 BP; 12193 A; 8355 C; 8658 G; 12815 T; 0 U; 40 Other;
XX
XX Query Match 75.2%; Score 18.8; DB 13; Length 42061;
XX Best Local Similarity 90.9%; Pred. No. 3.9e+02;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 AAAAAACACCACTGAGCT 22
XX ||||| ||||| ||||| |||||
XX 10151 AAAAAACACCACTGAGCT 10130
XX
XX RESULT 3
XX ACH01384
XX ID ACH01384 standard; DNA; 96898 BP.
XX
XX AC ACH01384;
XX
XX DT 22-APR-2004 (first entry)
XX
XX DE Human BIVM gene.
XX
XX KW Human; mouse; zebrafish; chicken; sea urchin; BIVM; immunostimulant;
XX basic immunoglobulin-like variable motif containing; cytostatic; cancer;
XX gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO2003089595-A2.
XX
XX PD 30-OCT-2003.
XX
XX PF 16-APR-2003; 2003WO-US011768.
XX
XX PR 16-APR-2002; 2002US-0373146P.
XX
XX PA (UYSF-) UNIV SOUTH FLORIDA.
XX

XX Litman GW, Hawke NA, Yoder JA, Eason DD;
XX
XX WPI; 2003-854106/79.
XX
XX New isolated or recombinant basic immunoglobulin-like variable motif-
XX containing (BIVM) polynucleotides and polypeptides, useful for inducing
XX an immune response or protective immunity against cancer, and for
XX detecting Giardia.
XX
XX Claim 1; Page 68-123; Opp; English.
XX
XX The present invention relates to isolated forms of basic, immunoglobulin-
XX like variable motif-containing (BIVM) genes and encoded proteins. These
XX can be used to induce an immune response or protective immunity against
XX cancer. The present sequence is a coding sequence fragment shown in the
XX exemplification of the invention
XX
XX Sequence 96898 BP; 27199 A; 18553 C; 20089 G; 31056 T; 0 U; 1 Other;
XX
XX Query Match 75.2%; Score 18.8; DB 10; Length 96898;
XX Best Local Similarity 90.9%; Pred. No. 4.4e+02;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 AAAAAACACCACTGAGCT 22
XX ||||| ||||| ||||| |||||
XX 5945 AAAAAACACCACTGAGCT 5966
XX
XX Db
XX
XX RESULT 4
XX ABZ09104/C
XX ID ABZ09104 standard; DNA; 90 BP.
XX
XX AC ABZ09104;
XX
XX DT 16-JAN-2003 (first entry)
XX
XX DE Human oligonucleotide SEQ ID 264.
XX
XX KW Human; tumour suppressor; virucide; cytostatic; nootropic;
XX neuroprotective; neuroleptic; gene therapy; tumour suppression;
XX tumour reversion; apoptosis; viral resistance; viral infection;
XX cell degeneration; Alzheimer's disease; schizophrenia; cancer; ds.
XX
XX OS Homo sapiens.
XX
XX PN FR2822475-A1.
XX
XX PD 27-SEP-2002.
XX
XX PF 20-MAR-2002; 2002FR-00003459.
XX
XX PR 13-FEB-2001; 2001FR-00001925.
XX
XX PA (MOLE-) MOLECULAR ENGINES LAB SA.
XX
XX PI Teلمان A, Anson R, Tuijnder M, Susini L;
XX
XX DR WPI; 2003-032204/03.
XX
XX PT New human nucleic acid, useful for diagnosis, prognosis and treatment,
XX e.g. of tumors, also related vectors, transformed cell, polypeptides and
XX antibodies.
XX
XX Disclosure; Page 78; 189pp; French.
XX
XX The present invention relates to human oligonucleotides (ABZ08841-
XX ABZ09860). The expression of the oligonucleotides is implicated in tumour
XX suppression or reversion, apoptosis and/or viral resistance. The
XX oligonucleotides are useful for preventing and/or treating viral
XX infection, tumour development and cell degeneration (e.g. Alzheimer's
XX disease and schizophrenia), especially cancer
XX
```

```

SQ Sequence 90 BP; 30 A; 16 C; 13 G; 31 T; 0 U; 0 Other;
  Query Match          74.4%; Score 18.6; DB 8; Length 90;
  Best Local Similarity 84.0%; Pred. No. 1.8e+02;
  Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACAACCACTGAGCTGGG 25
  |||||
DB 43 AAAAAACAACCACTGGAATGTG 19

RESULT 5
ABZ78557/c
ID ABZ78557 standard; DNA; 90 BP.
XX AC ABZ78557;
XX DT 24-APR-2003 (first entry)
XX XX Tumour suppression-related sequence, SEQ ID 264.
XX DE Cytostatic; virucide; apoptotic; gene therapy; tumour suppression;
KW tumour reversion; apoptosis; virus resistance; viral infection; tumour;
KW cell degenerative disease; ds.
XX OS Unidentified.
XX XX WO200264731-A2.
XX PN 22-AUG-2002.
XX PD 13-FEB-2002; 2002WO-FR000543.
XX PF 13-FEB-2001; 2001FR-00001925.
XX PR (MOLE-) MOLECULAR ENGINES LAB.
XX PA Telerman A, Anson R, Tuijnder M, Susini L;
PI WPI; 2003-058286/05.
XX DR New nucleic acid encoding a translationally controlled tumor protein,
PT useful for treating, preventing and diagnosing viral, tumor or
PT degenerative diseases.
XX PS Disclosure; Page; 45pp; French.
XX CC The present invention relates to novel nucleic acid sequences (ABZ78294-
CC ABZ79313), which are involved in the molecular pathways of tumour
CC suppression, tumour reversion, apoptosis and/or virus resistance. The
CC sequences are also useful for treatment or prevention of viral, tumour
CC and cell degenerative diseases, and also for diagnosis and prognosis of
CC these diseases. Note: The sequence data for this patent is not
CC represented in the printed specification but is based on sequence
CC information supplied by the European Patent Office
XX XX
SQ Sequence 90 BP; 30 A; 16 C; 13 G; 31 T; 0 U; 0 Other;
  Query Match          74.4%; Score 18.6; DB 10; Length 90;
  Best Local Similarity 84.0%; Pred. No. 1.8e+02;
  Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACAACCACTGAGCTGGG 25
  |||||
DB 43 AAAAAACAACCACTGGAATGTG 19

RESULT 6
AAT24655
ID AAT24655 standard; cDNA to mRNA; 93 BP.
XX AC AAT24655;
XX XX

SQ Sequence 93 BP; 34 A; 13 C; 16 G; 30 T; 0 U; 0 Other;
  Query Match          74.4%; Score 18.6; DB 2; Length 93;
  Best Local Similarity 84.0%; Pred. No. 1.8e+02;
  Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACAACCACTGAGCTGGG 25
  |||||
DB 48 AAAAAACAACCACTGGAATGTG 72

RESULT 7
ADD47465
ID ADD47465 standard; DNA; 382 BP.
XX AC ADD47465;
XX XX
DT 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX XX
DE Rat gene AA852046, SEQ ID NO 13160.
XX XX
KW Rat; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX XX
OS Rattus norvegicus.
```

OS Unidentified.  
XX WO2003016475-A2.  
PN  
XX  
XX  
XX 27-FEB-2003.  
PD  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
PR  
XX 01-NOV-2001; 2001US-0346382P.  
PR  
XX 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
PA  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
DR GENBANK; AAB52046.  
DR  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
PT  
XX  
XX Example 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat DNA (described in Table 3 of  
CC the specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 382 BP; 144 A; 102 C; 69 G; 65 T; 0 U; 2 Other;  
SQ  
  
Query Match 74.4%; Score 18.6; DB 10; Length 382;  
Best Local Similarity 84.0%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 AAAAAACACACACACACACGCTGGG 25  
Db 101 AACCAACACACCGCACCTGACGGGG 125  
  
RESULT 8  
ABA08904  
ID ABA08904 standard; cDNA; 540 BP.  
XX  
XX ABA08904;  
XX  
XX 11-JAN-2002 (first entry)  
XX

DE Human olfactory receptor homologue-encoding cDNA, SEQ ID NO:680.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiaschmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antiulcer; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157188-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US003800.  
PF  
XX  
XX 03-FEB-2000; 2000US-00496914.  
PR  
XX 27-APR-2000; 2000US-00560875.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-457740/49.  
DR P-PSDB; ABB11660.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing, treating  
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
PT and cancer.  
PT  
XX  
XX Claim 1; Page 638; 1963pp; English.  
PS  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,

CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention

XX  
SQ Sequence 540 BP; 139 A; 132 C; 160 G; 109 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 540;

Best Local Similarity 84.0%; Pred. No. 2.4e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 323 AAAAAACCACTGCGGTG 347

RESULT 9

AD121508

ID AD121508 standard; cDNA; 540 BP.

XX

AC AD121508;

XX

DT 15-APR-2004 (first entry)

XX

DE Novel human expressed sequence tag, EST #207.

XX

KW forensic; nutritional source; damaged tissue; diseased tissue;

myeloid cell disorder; lymphoid cell disorder;

bone cartilage tissue growth; tendon tissue growth;

ligament tissue growth; nerve tissue growth; regeneration; wound healing;

tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;

ss; Gene; EST; expressed sequence tag.

XX

OS Homo sapiens.

XX

PN WO2003025148-A2.

XX

PD 27-MAR-2003.

XX

PF 19-SEP-2002; 2002WO-US029964.

XX

PR 19-SEP-2001; 2001US-0323739P.

PR

PR 13-SEP-2002; 2002US-00323739.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;

PI Haley-Vicente D;

XX

DR WPI; 2003-354603/33.

DR

P-PSDB; AD121728.

XX

PT New polynucleotides and secreted proteins, useful for treating myeloid or  
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
PT tissue growth or regeneration, in wound healing, and in tissue repair and  
PT replacement.

XX

PS Example 2; SEQ ID NO 759; 156pp; English.

XX

CC The invention relates to an isolated polynucleotide encoding a

polypeptide with biological activity. The polynucleotides and

polypeptides are useful in diagnostics, forensics, gene mapping,

CC identification of mutations responsible for genetic disorders and other

traits, to assess biodiversity, as nutritional sources or supplements.

CC The polynucleotides may also be used as molecular weight markers,

CC chromosome markers or map related gene positions, or as an antigen to

raise anti-DNA antibodies or elicit immune response. The polypeptides are

CC useful for raising antibodies, as markers for tissues in which the

corresponding polypeptide is expressed, for re-engineering damaged or

CC diseased tissues, for treating myeloid or lymphoid cell disorders, in

CC bone cartilage, tendon, ligament and/or nerve tissue growth or

regeneration, in wound healing, in tissue repair and replacement, in

CC

CC healing of burns, incisions and ulcers, and in treating cancer. The  
CC present sequence represents a novel human expressed sequence tag, EST.

XX  
SQ Sequence 540 BP; 139 A; 132 C; 160 G; 109 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 10; Length 540;

Best Local Similarity 84.0%; Pred. No. 2.4e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 323 AAAAAACCACTGCGGTG 347

RESULT 10

ADS98314

ID ADS98314 standard; DNA; 540 BP.

XX

AC ADS98314;

XX

DT 30-DEC-2004 (first entry)

XX

DE Protein factor discovery related isolated human contig DNA, SEQ ID 578.

XX

KW antiinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;

leukaemia; nervous system disorder; infection; gene; ds.

XX

OS Homo sapiens.

XX

PN WO2004087874-A2.

XX

PD 14-OCT-2004.

XX

PF 24-MAR-2004; 2004WO-US009202.

XX

PR 28-MAR-2003; 2003US-0458824P.

XX

PA (NUVE-) NUVELO INC.

PA (DRNA/) DRMANAC R T.

XX

PI Tang YT, Zhou P, Wang J, Wang ZW, Hu T;

XX

WPI; 2004-737686/72.

DR

P-PSDB; ADS98654.

XX

PT New polynucleotides encoding a polypeptide with biological activity,  
PT useful for treating inflammation, leukemias, nervous system disorders, or  
PT infections.

XX

PS Example 2; SEQ ID NO 578; 253pp; English.

XX

CC The invention relates to a novel isolated polynucleotide comprising any  
of the 235 nucleotide sequences described in the specification. The  
CC invention further comprises: an isolated polynucleotide encoding a

polypeptide with biological activity, where the polynucleotide hybridizes

to one of the 235 novel polynucleotides under stringent hybridization

conditions, or having greater than about 99% sequence identity with the

novel polynucleotide; a vector comprising a novel polynucleotide; an

expression vector comprising the novel polynucleotide; a host cell

genetically engineered to comprise the novel polynucleotide, which can be

operatively associated with a regulatory sequence that modulates

expression of the polynucleotide in the host cell; an isolated

polypeptide encoded by the novel polynucleotide, or a polynucleotide

hybridizing under stringent conditions to the novel polynucleotide; a

composition comprising the polypeptide and a carrier; an antibody

directed against the polypeptide; a method for detecting the novel

polynucleotide in a sample; a method for detecting the polypeptide in a

sample; a method for identifying a compound that binds to the polypeptide

; a method for producing the polypeptide; an isolated polypeptide

comprising any of the 235 amino acid sequences described in the

specification; and a collection of polynucleotides comprising of at least

one of the polynucleotides cited above. The polypeptides and

polynucleotides of the invention have antiinflammatory, cytostatic, and



CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,  
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,  
 CC Alzheimer's disease and as a contraceptive  
 XX  
 SQ Sequence 762 BP; 244 A; 138 C; 187 G; 178 T; 0 U; 15 Other;  
 Query Match 74.4%; Score 18.6; DB 6; Length 762;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||  
 Db 711 AAAAAACACCACTGAGCTGG 735  
 |||||  
 RESULT 13  
 ABZ42980/c  
 ID ABZ42980 standard; DNA; 777 BP.  
 XX  
 AC ABZ42980;  
 XX  
 DT 06-MAR-2003 (first entry)  
 XX  
 DE Human GPCR polynucleotide SEQ ID NO 221.  
 XX  
 KW Human; GPCR; G protein coupled receptor; signal transduction; olfactory;  
 KW drug development; gustatory; taste; fragrance; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200216548-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 30-JUL-2001; 2001WO-IB001446.  
 XX  
 PR 04-AUG-2000; 2000JP-00237818.  
 PR 13-FEB-2001; 2001JP-00034433.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Haga T, Takeda S, Mitaku S;  
 XX  
 DR WPI; 2002-304118/34.  
 DR P-PSDB; ABP95706.  
 XX  
 PT Database global search for G protein-coupled receptors, proteins and  
 PT encoded genes for studying in vivo signal transduction mechanism and  
 PT identifying targets for drug development.  
 XX  
 PS Claim 9; SEQ ID NO 221; 97pp + Sequence Listing; Japanese.  
 XX  
 CC The invention relates to a method for screening G protein-coupled  
 CC receptor (GPCR) genes (ABZ42980-ABZ43216) and/or GPCR proteins (ABP95596-  
 CC ABP95942) by extracting open-reading frames containing 6-8 transmembrane  
 CC domains with 250-1000 amino acid residues to give a gene homologous with  
 CC a known GPCR gene. The receptor proteins and encoded genes are useful for  
 CC studying in vivo signal transduction mechanism and identifying targets  
 CC for drug development e.g. based on olfactory and gustatory receptors in  
 CC form of agonists and antagonists by screening intrinsic and extrinsic  
 CC ligands as bitter taste inhibitors, taste enhancers and fragrance  
 CC improvers. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 777 BP; 145 A; 241 C; 194 G; 197 T; 0 U; 0 Other;  
 Query Match 74.4%; Score 18.6; DB 6; Length 777;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||

Db 544 AAAAAACCACTGTGGGTG 520  
 RESULT 14  
 ACL31584/c  
 ID ACL31584 standard; cDNA; 812 BP.  
 XX  
 AC ACL31584;  
 XX  
 DT 02-JUN-2005 (first entry)  
 XX  
 DE Rice abiotic stress response related polynucleotide SEQ ID NO:10147.  
 XX  
 KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;  
 KW agriculture.  
 XX  
 OS Oryza sativa.  
 XX  
 PN WO2003008540-A2.  
 XX  
 PD 30-JAN-2003.  
 XX  
 PF 21-JUN-2002; 2002WO-US019668.  
 XX  
 PR 22-JUN-2001; 2001US-0300112P.  
 PR 24-AUG-2001; 2001US-0314662P.  
 PR 26-SEP-2001; 2001US-0325277P.  
 PR 21-NOV-2001; 2001US-0332132P.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;  
 PI Moughamer T, Provart N, Ricke D, Zhu T;  
 XX  
 DR WPI; 2003-248011/24.  
 XX  
 PT New stress-responsive nucleic acid, useful for altering the  
 PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold  
 PT stress, salt stress or osmotic stress.  
 XX  
 PS Disclosure; SEQ ID NO 10147; 89pp; English.  
 XX  
 CC The invention relates to novel abiotic stress responsive polynucleotides  
 CC and polypeptides. Also disclosed are vectors, expression cassettes, host  
 CC cells, and plants containing such polynucleotides. Also disclosed are  
 CC methods for using the polynucleotides and polypeptides to alter the  
 CC responsiveness of a plant to abiotic stress. The invention is useful in  
 CC agriculture. The nucleic acid is useful for determining whether a test  
 CC plant has been exposed to an abiotic stress condition. It is also useful  
 CC for selecting an agent that alters abiotic stress regulated  
 CC polynucleotide expression in a plant cell, and to identify a homolog or  
 CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid  
 CC molecule and the polypeptide encoded by it are useful in altering the  
 CC responsiveness of a plant to an abiotic stress, such as cold stress, salt  
 CC stress, osmotic stress or any of their combinations. The present sequence  
 CC is used in the exemplification of the invention  
 XX  
 SQ Sequence 812 BP; 171 A; 212 C; 225 G; 204 T; 0 U; 0 Other;  
 Query Match 74.4%; Score 18.6; DB 11; Length 812;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACCACTGAGCTGGG 25  
 |||||  
 Db 509 AAGAATCAACCACTTACTGAGCTGGG 485  
 |||||  
 RESULT 15  
 AAH32071/c  
 ID AAH32071 standard; DNA; 932 BP.  
 XX  
 AC AAH32071;



XX 30-JUL-2001 (first entry)  
 XX Human olfactory receptor polynucleotide, SEQ ID NO: 644.  
 DE  
 XX  
 XX Human; olfactory receptor; OR; primary scent determination;  
 KW secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200127158-A2.  
 XX  
 XX 19-APR-2001.  
 XX  
 XX 06-OCT-2000; 2000WO-US027582.  
 XX  
 XX 08-OCT-1999; 99US-0158615P.  
 XX  
 XX 24-FEB-2000; 2000US-0184809P.  
 XX  
 XX (DIGI-) DIGISCENTS.  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 PA  
 XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
 XX  
 XX WPI; 2001-290713/30.  
 DR  
 XX New polynucleotides which encode polypeptides involved in olfactory  
 PT sensation for identifying olfactory agonists and antagonists.  
 PT  
 XX Claim 8; Page 439; 1857pp; English.  
 XX  
 XX The present sequence is one of a number of isolated polynucleotides which  
 CC encode polypeptides involved in olfactory sensation. The polynucleotides  
 CC can be used in screening for olfactory agonists and antagonists. The  
 CC methods allow for the determination of primary scents and the  
 CC identification of the odour receptors used to detect these primary  
 CC scents. The methods also enable determination of secondary scents and the  
 CC identification of combinations of odour receptors that are involved in  
 CC detecting such secondary scents. This enables the construction of a scent  
 CC representation (also called a scent fingerprint or scent profile), which  
 CC may be used to re-create and edit scents. Libraries of olfactory  
 CC receptors are useful for determining the interaction pattern of a  
 CC composition with the receptors, and can be used for determining  
 CC differences in the olfactory faculties of different individuals  
 XX  
 XX Sequence 932 BP; 179 A; 284 C; 236 G; 233 T; 0 U; 0 Other;  
 SQ  
 Query Match 74.4%; Score 18.6; DB 4; Length 932;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACACCACCTGAGCTGGG 25  
 DB 744 AAAAAACACACCACCTGAGCTGGG 720  
 RESULT 16  
 ABX34731/c  
 ID ABX34731 standard; cDNA; 950 BP.  
 XX  
 XX ABX34731;  
 AC  
 XX 13-FEB-2003 (first entry)  
 DT  
 XX Human mddt cDNA SEQ ID 292.  
 DE  
 XX MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cycostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;

KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis; gene; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200279449-A2.  
 FN  
 XX 10-OCT-2002.  
 PD  
 XX 27-MAR-2002; 2002WO-US009944.  
 XX  
 XX 28-MAR-2001; 2001US-0279619P.  
 XX  
 XX 29-MAR-2001; 2001US-0280067P.  
 PR  
 XX 29-MAR-2001; 2001US-0280068P.  
 PR  
 XX 16-MAY-2001; 2001US-0291280P.  
 PR  
 XX 17-MAY-2001; 2001US-0291829P.  
 PR  
 XX 17-MAY-2001; 2001US-0291849P.  
 PR  
 XX 19-JUN-2001; 2001US-0299428P.  
 PR  
 XX 20-JUN-2001; 2001US-0299776P.  
 PR  
 XX 20-JUN-2001; 2001US-0300001P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Paralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RV, Urashka ME;  
 XX  
 XX WPI; 2003-058431/05.  
 DR  
 XX P-PSDB; ABU11741.  
 XX  
 XX New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis.  
 XX  
 XX Claim 1; SEQ ID NO 292; 339pp + Sequence Listing; English.  
 XX  
 XX This invention describes a novel disease detection and treatment molecule  
 CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
 CC osteopathic, cycostatic, anti-HIV, haemostatic, nephrotropic,  
 CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or  
 CC hepatitis. ABX34440-ABX34835 encode the MDDT polypeptides represented in  
 CC ABU1450-ABU11845, described in the disclosure of the invention. NOTE:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 950 BP; 173 A; 300 C; 239 G; 238 T; 0 U; 0 Other;  
 SQ  
 Query Match 74.4%; Score 18.6; DB 8; Length 950;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACACCACCTGAGCTGGG 25  
 DB 690 AAAAAACACACCACCTGAGCTGGG 666  
 RESULT 17  
 ABZ77938/c  
 ID ABZ77938 standard; DNA; 951 BP.  
 XX  
 XX AC ABZ77938;

XX 17-APR-2003 (first entry)  
 DT Human G protein coupled receptor coding sequence SEQ ID 133.  
 XX  
 DE Human; gene; anorectic; antiasthmatic; antidiabetic; hypotensive;  
 XX antiparkinsonian; nootropic; neuroprotective; tranquiliser;  
 KW antirheumatic; antiinflammatory; osteopathic; cardiant; neuroleptic;  
 KW antithratic; gene therapy; olfactory G protein-coupled receptor; GPCR;  
 KW infection; obesity; diabetes; hypertension; malnutrition;  
 KW Parkinson's disease; Alzheimer's disease; Korsakoff's psychosis; anxiety;  
 KW rheumatoid arthritis; chronic obstructive pulmonary disease;  
 KW osteoporosis; asthma; myocardial infarction; schizophrenia;  
 KW osteoarthritis; db.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO2003000735-A2.  
 PN  
 XX 03-JAN-2003.  
 PD  
 XX 24-JUN-2002; 2002WO-IB002481.  
 XX  
 XX 26-JUN-2001; 2001US-0301095P.  
 PR  
 XX 06-NOV-2001; 2001US-0332758P.  
 XX  
 XX (DECO-) DECODE GENETICS EHF.  
 PA  
 XX Martinez RAM, Sigurdsson GT;  
 XX  
 XX WPI; 2003-175284/17.  
 DR  
 XX P-PSDB; ABR01637.  
 XX  
 XX New olfactory G protein-coupled receptor gene nucleic acid and  
 PT polypeptide, useful for diagnosing or treating a disease or condition  
 PT associated with GPCR, e.g. Obesity, diabetes, hypertension, malnutrition  
 PT or Alzheimer's disease.  
 XX  
 XX Claim 1; Page 87; 383pp; English.  
 PS  
 XX The present invention relates to novel human olfactory G protein-coupled  
 CC receptors (GPCR) and their coding sequences (ABZ77872-ABZ77986 and  
 CC ABR01571-ABR01685). The GPCRs and coding sequences are useful for  
 CC diagnosing or treating a disease or condition associated with GPCR, e.g.  
 CC infections, obesity, diabetes, hypertension, malnutrition, Parkinson's  
 CC disease, Alzheimer's disease, Korsakoff's psychosis, anxiety, rheumatoid  
 CC arthritis, chronic obstructive pulmonary disease, osteoporosis, asthma,  
 CC myocardial infarction, schizophrenia, or osteoarthritis  
 XX  
 XX Sequence 951 BP; 177 A; 298 C; 235 G; 241 T; 0 U; 0 Other;  
 SQ  
 Query Match 74.4%; Score 18.6; DB 10; Length 951;  
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACCACCTGAGCTGG 25  
 |||||  
 Db 766 AAAAAACACCACCTGTCGGTG 742  
 |||||  
 RESULT 18  
 AAS15907/c  
 ID AAS15907 standard; cDNA; 1065 BP.  
 XX  
 AC AAS15907;  
 XX  
 DT 25-JAN-2002 (first entry)  
 XX  
 XX DNA encoding G-protein coupled receptor (GCREC) #11.  
 DE  
 XX G-protein coupled receptor; GCREC; vaccine; gene therapy;  
 KW cell proliferation disorder; cancer; arteriosclerosis;  
 KW neurological disorder; epilepsy; stroke; cardiovascular disorder;  
 KW

KW hypertension; ischaemic heart disease; gastrointestinal disorder;  
 KW anorexia; peptic ulcer; autoimmune disorder; inflammatory disorder;  
 KW diabetes; osteoporosis; psoriasis; metabolic disorder; obesity;  
 KW schizophrenia disorder; neuroskeletal disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 CDS 1..1065  
 FT /\*tag= a  
 FT /product= "GCREC 11"  
 FT /note= "G-protein coupled receptor 11"  
 XX  
 PN WO200166742-A2.  
 XX  
 XX 13-SEP-2001.  
 XX  
 XX 01-MAR-2001; 2001WO-US006814.  
 XX  
 XX 03-MAR-2000; 2000US-0186854P.  
 PR  
 XX 10-MAR-2000; 2000US-0188384P.  
 PR  
 XX 17-MAR-2000; 2000US-0190453P.  
 PR  
 XX 20-MAR-2000; 2000US-0190730P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Lal P, Tang YT, Patterson C, Yao MG, Shih LL, Tribouley CM;  
 PI Lu DAM, Yue H, Khan FA, Policky JL, Au-Young J, Yang J, Harland L;  
 PI Walsh RT, Lo TP, Borowsky ML;  
 XX  
 XX WPI; 2001-656776/75.  
 DR  
 XX P-PSDB; AAU10310.  
 XX  
 XX Novel G-protein coupled receptor polypeptides, for treating and  
 PT preventing autoimmune/inflammatory disorders, neurological disorders,  
 PT cell proliferative disorders, cardiovascular disorders and viral  
 PT infections.  
 XX  
 XX Claim 5; Page 136-137; 141pp; English.

XX The invention describes a novel isolated polypeptide, selected from a  
 CC group of 21 G-protein coupled receptor polypeptides (GCREC) and useful in  
 CC vaccines and gene therapy. The polypeptide (I) is useful for screening  
 CC for agonist or antagonist of (I), compounds specifically binding to (I),  
 CC or compounds that modulate the activity of (I). The polynucleotide  
 CC encoding (I) is useful for screening a compound for effectiveness in  
 CC altering expression of a target polynucleotide comprising (II), by  
 CC exposing a sample comprising the target polynucleotide to a compound,  
 CC detecting altered expression of the target polynucleotide, and comparing  
 CC the expression of the target polynucleotide in the presence of varying  
 CC amounts of compound and in the absence of the compound. (I) and (II) are  
 CC useful for diagnosis, treatment and prevention of cell proliferative  
 CC disorders (e.g. cancers, arteriosclerosis, atherosclerosis), neurological  
 CC disorders (e.g. epilepsy, stroke, schizophrenic disorders and  
 CC neuroskeletal disorders), cardiovascular disorders (e.g. hypertension,  
 CC ischaemic heart disease), gastrointestinal disorders (e.g. anorexia,  
 CC peptic ulcer), autoimmune/inflammatory disorders (e.g. diabetes mellitus,  
 CC osteoporosis, psoriasis), and metabolic disorders such as obesity.  
 CC Furthermore, the polynucleotide is useful; as primers for detecting  
 CC single nucleotide polymorphisms; as elements in microarray, to monitor or  
 CC measure protein-protein interactions, drug-target interactions, and gene  
 CC expression profiles; to generate a transcript image of a tissue or cell  
 CC type, and to generate hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequence. This sequence encodes G-protein  
 CC coupled receptor 11, one of 21 GCREC proteins described in the method of  
 CC the invention  
 XX  
 XX Sequence 1065 BP; 214 A; 327 C; 271 G; 253 T; 0 U; 0 Other;  
 SQ

Query Match 74.4%; Score 18.6; DB 5; Length 1065;  
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25  
DB 832 AAAAAACACCACTGAGCTGGG 808

RESULT 19  
ABK40204/c  
ID ABK40204 standard; cDNA; 1069 BP.  
XX AC ABK40204;  
XX 21-MAY-2002 (first entry)  
XX DE Human G protein-coupled receptor (GPCR) 9 cDNA #2.  
XX Human; G protein-coupled receptor; GPCR; cell signalling processing; ss;  
KW cardiomyopathy; atherosclerosis; developmental disease; immune disease;  
KW non insulin-dependent diabetes mellitus; NIDDM1; bacterial infection;  
KW fungal infection; protozoal infection; viral infection; adenocarcinoma;  
KW human immunodeficiency virus; HIV; cancer; lymphoma; anorexia; asthma;  
KW neurodegenerative disorder; Parkinson's disease; acute heart failure;  
KW hypotension; hypertension; osteoporosis; Crohn's disease; allergy; ulcer;  
KW multiple sclerosis; osteodystrophy; angina pectoris; schizophrenia; gene;  
KW myocardial infarction; neurological disorder; anxiety; manic depression;  
KW delirium; dementia; Huntington's disease; Alzheimer's disease; obesity;  
KW Tourette's syndrome; metabolic disorder.  
XX OS Homo sapiens.  
XX WO200202637-A2.  
XX 10-JAN-2002.  
XX 02-JUL-2001; 2001WO-US021174.  
XX 30-JUN-2000; 2000US-0215483P.  
XX 30-JUN-2000; 2000US-0215495P.  
XX 08-JUL-2000; 2000US-0216117P.  
XX 25-JUL-2000; 2000US-0220589P.  
XX 02-NOV-2000; 2000US-0245294P.  
XX 10-JAN-2001; 2001US-0260729P.  
XX 10-JAN-2001; 2001US-0260851P.  
XX 26-FEB-2001; 2001US-0271673P.  
XX (CURA-) CURAGEN CORP.  
XX Spaderna BK, Padigar M, Spytek KA, Casman S, Rastelli L;  
PI Mayankar U, Tchernev V;  
XX WPI; 2002-164524/21.  
XX P-PSDB; AAU85873.  
XX Novel polypeptides and nucleic acids of G protein coupled receptor useful  
PT for diagnosing and treating diseases e.g. anorexia, asthma, Parkinson's  
PT disease, acute heart failure and infections.  
XX Claim 9; Page 58; 202pp; English.  
XX The invention relates to human G protein coupled receptor (GPCR)  
CC polypeptides and the polynucleotides encoding them. GPCR polypeptides are  
CC useful for treating or preventing a GPCR-associated disorder which is  
CC related to cell signalling processing and metabolic pathway modulation,  
CC which includes cardiomyopathy and atherosclerosis in a human. The  
CC sequences are useful in treatment of patients suffering from  
CC developmental diseases, immune diseases, non insulin-dependent diabetes  
CC mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections  
CC (e.g. infections caused by HIV), cancer (e.g. adenocarcinoma, lymphoma),  
CC anorexia, asthma, neurodegenerative disorders (e.g. Parkinson's disease),  
CC acute heart failure, hypertension, osteoporosis, Crohn's  
CC disease, multiple sclerosis, osteodystrophy, angina pectoris, myocardial  
CC infarction, ulcers, allergies, neurological disorders including anxiety,  
CC schizophrenia, manic depression, delirium, dementia, Huntington's  
CC disease, Alzheimer's disease, Tourette's syndrome, metabolic disorders

CC and obesity. This sequence represents cDNA encoding a GPCR polypeptide of  
CC the invention  
XX Sequence 1069 BP; 210 A; 315 C; 274 G; 270 T; 0 U; 0 Other;  
SQ Query Match 74.4%; Score 18.6; DB 6; Length 1069;  
Best Local Similarity 84.0%; Pred. No. 2.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACCACTGAGCTGGG 25  
DB 778 AAAAAACCACTGAGCTGGG 754

RESULT 20  
ADS97926/c  
ID ADS97926 standard; DNA; 1232 BP.  
XX ADS97926;  
XX 30-DEC-2004 (first entry)  
XX Protein factor discovery related isolated human DNA, SEQ ID No 190.  
XX antinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;  
KW leukaemia; nervous system disorder; infection; gene; ds.  
XX OS Homo sapiens.  
XX WO2004087874-A2.  
XX 14-OCT-2004.  
XX 24-MAR-2004; 2004WO-US009202.  
XX 28-MAR-2003; 2003US-0458824P.  
XX (NUVE-) NUVELO INC.  
XX (DRMA/) DRMANAC R T.  
XX Tang YT, Zhou P, Wang J, Wang ZW, Hu T;  
XX WPI; 2004-737686/72.  
XX P-PSDB; ADS98161.  
XX New polynucleotides encoding a polypeptide with biological activity,  
PT useful for treating inflammation, leukemias, nervous system disorders, or  
PT infections.  
XX Claim 1; SEQ ID NO 190; 253pp; English.  
XX The invention relates to a novel isolated polynucleotide comprising any  
CC of the 235 nucleotide sequences described in the specification. The  
CC invention further comprises: an isolated polynucleotide encoding a  
CC polypeptide with biological activity, where the polynucleotide hybridizes  
CC to one of the 235 novel polynucleotides under stringent hybridization  
CC conditions, or having greater than about 99% sequence identity with the  
CC novel polynucleotide; a vector comprising a novel polynucleotide; an  
CC expression vector comprising the novel polynucleotide; a host cell  
CC genetically engineered to comprise the novel polynucleotide, which can be  
CC operatively associated with a regulatory sequence that modulates  
CC expression of the polynucleotide in the host cell; an isolated  
CC polypeptide encoded by the novel polynucleotide, or a polynucleotide  
CC hybridizing under stringent conditions to the novel polynucleotide; a  
CC competition comprising the polypeptide and a carrier; an antibody  
CC directed against the polypeptide; a method for detecting the novel  
CC polynucleotide in a sample; a method for detecting the polypeptide in a  
CC sample; a method for identifying a compound that binds to the polypeptide  
CC ; a method for producing the polypeptide; an isolated polypeptide  
CC comprising any of the 235 amino acid sequences described in the  
CC specification; and a collection of polynucleotides comprising of at least  
CC one of the polynucleotides cited above. The polypeptides and  
CC polynucleotides of the invention have antinflammatory, cytostatic, and

CC antimicrobial activities. The novel polynucleotide may be used to treat  
 CC disorders by gene therapy. The polypeptides and polynucleotides are  
 CC useful for treating inflammation, leukaemias, nervous system disorders,  
 CC or infections. This sequence represents one of the 235 novel isolated  
 CC polynucleotides of the invention.

XX  
 SQ Sequence 1232 BP; 252 A; 358 C; 315 G; 307 T; 0 U; 0 Other;  
 Query Match 74.4%; Score 18.6; DB 13; Length 1232;  
 Best Local Similarity 84.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 4;

QY 1 AAAAAACACCACTGAGCTGGG 25

Db 916 AAAAAACCACTGCTGGGTG 892

# RESULT 21

ABN89112/c

ID ABN89112 standard; cDNA; 1329 BP.

XX AC ABN89112;

XX DT 28-AUG-2002 (first entry)

XX DE Human GPCR1 nucleotide sequence SEQ ID NO:1.

XX KW Human; GPCR; GPCR; G protein-coupled receptor; antiatherosclerotic;  
 KW anorectic; antibacterial; fungicide; protozoacide; virucide; analgesic;  
 KW cytotatic; immunomodulator; metabolic; antisthmatic; antiparkinsonian;  
 KW hypertensive; hypotensive; osteopathic; antinflammatory; anti-HIV;  
 KW antinfertility; neuroprotective; antianigmal; cardiant; antiulcer;  
 KW antiallergic; nootropic; tranquiliser; neuroleptic; antidepressant;  
 KW antimanic; anticonvulsant; haemostatic; immunosuppressive; hepatotropic;  
 KW cerebroprotective; dermatological; antiaddictive; gynaecological; cancer;  
 KW nephrotropic; gene therapy; vaccine; developmental disease; diabetes;  
 KW cardiomyopathy; atherosclerosis; neurodegenerative disorder;  
 KW autoimmune disorder; infectious disease; chromosome 9; gene; ss.

XX OS Homo sapiens.

XX PN WO200226985-A2.

XX XX 04-APR-2002.

XX PF 28-SEP-2001; 2001WO-US030552.

XX PR 28-SEP-2000; 2000US-0236284P.

XX PR 28-SEP-2000; 2000US-0236286P.

XX PR 03-OCT-2000; 2000US-0237581P.

XX PR 06-OCT-2000; 2000US-0238735P.

XX PR 16-OCT-2000; 2000US-0240736P.

XX PR 05-JAN-2001; 2001US-0260019P.

XX PR 08-JAN-2001; 2001US-0260338P.

XX PR 17-JAN-2001; 2001US-0262156P.

XX PR 18-JAN-2001; 2001US-0262498P.

XX PR 19-JAN-2001; 2001US-0263133P.

XX PR 24-JAN-2001; 2001US-0263691P.

XX PR 02-FEB-2001; 2001US-0266109P.

XX PR 26-FEB-2001; 2001US-0271634P.

XX PR 27-SEP-2001; 2001US-00965422.

XX (CURA-) CURAGEN CORP.

XX PI Spytek KA, Casman S, Padigar M, Dickson K, Vernet C;

PI Spaderna SK, Shenoy S, Gerlach V, Ellerman K, Edinger S;

PI Macdougall JR, Smithson G, Li L, Malyankar UM, Taylor S, Gunther E;

PI Tchernev VT;

XX WPI; 2002-499868/53.

XX P-PSDB; ABB81440.

XX Novel G-protein coupled-receptor polypeptides and nucleic acids for

PT diagnosing, treating cardiomyopathy, atherosclerosis, cancer,  
 PT neurodegenerative, autoimmune disorders, infectious diseases and  
 PT diabetes.

XX Claim 8; Page 11; 213pp; English.

XX ABN89112 to ABN89130 encoding the human G protein-coupled receptor  
 CC (GPCR) proteins given in ABB81440 to ABB81458. GPCR proteins can have  
 CC antiatherosclerotic, anorectic, antibacterial, fungicide, protozoacide,  
 CC virucide, analgesic, cytotatic, immunomodulator, metabolic, hypotensive,  
 CC antisthmatic, antiparkinsonian, hypertensive, osteopathic, anti-HIV,  
 CC antinflammatory, antinfertility, neuroprotective, antianigmal,  
 CC cardiant, antiulcer, antiallergic, nootropic, tranquiliser, hepatotropic,  
 CC antidepressant, antimanic, anticonvulsant, haemostatic, antiaddictive,  
 CC immunosuppressive, cerebroprotective, dermatological, antiaddictive,  
 CC nephrotropic and gynaecological activities, and can be used in vaccines  
 CC and gene therapy. GPCR proteins and polynucleotides can be used for  
 CC treating or preventing a GPCR-associated disorder such as  
 CC cardiomyopathy, atherosclerosis, a disorder related to cell signal  
 CC processing and metabolic pathway modulation, in a human. GPCR proteins  
 CC and polynucleotides can also be used in the manufacture of a medicament  
 CC for preventing or treating disorders or syndromes including developmental  
 CC diseases, cardiomyopathy, atherosclerosis, cancer, neurodegenerative  
 CC disorders, autoimmune disorders, infectious diseases and diabetes.  
 CC ABN89131 to ABN89187 represent PCR primers and probes used in the  
 CC exemplification of the present invention

XX SQ Sequence 1329 BP; 275 A; 397 C; 326 G; 331 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 1329;

Best Local Similarity 84.0%; Pred. No. 2.8e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

Db 748 AAAAAACCACTGCTGGGTG 724

# RESULT 22

ABA95482

ID ABA95482 standard; cDNA; 1344 BP.

XX AC ABA95482;

XX DT 12-MAR-2002 (first entry)

XX DE Human laminin protein B2 chain 13 coding sequence.

XX KW Human; laminin protein B2 chain 13; cytotatic; immunomodulatory;  
 KW antinflammatory; neuroprotective; anti-HIV; neurological disease;  
 KW phlogosis; cancer; allopasia; HIV infection; immunological disease;  
 KW inflammatory disease; inflammation; ss.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 171..530

XX /\*tag= a

XX /product= "Human laminin protein B2 chain 13"

XX WO200190165-A1.

XX PD 29-NOV-2001.

XX PF 28-APR-2001; 2001WO-CN000641.

XX PR 29-APR-2000; 2000CN-00115524.

XX PA (SHAN-) SHANGHAI BIONDOWN GENE DEV INC.

XX PI Mao Y, Xie Y;

XX WPI; 2002-106178/14.

DR P-PSDB; AAM48188.  
XX HOMO laminin protein B2 chain 13 polynucleotide and polypeptide, useful  
PT in diagnosis and treatment of neurological diseases, phlogosis, cancer,  
XX immunological diseases and various inflammatory diseases.  
XX  
XX Claim 6; Page 30-31; 39pp; Chinese.  
XX  
XX The present sequence is the coding sequence for human laminin protein B2  
CC chain 13. The protein and its coding sequence are useful in the diagnosis  
CC and treatment of neurological diseases, phlogosis, cancer, allopiaasia,  
CC HIV infection, immunological diseases and various inflammatory diseases  
XX  
XX Sequence 1344 BP; 380 A; 293 C; 301 G; 370 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 1344;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1295 AAAAAACACACCACTGAGCTGGG 1319

RESULT 23  
ABK92484  
ID ABK92484 standard; DNA; 1450 BP.  
XX  
AC ABK92484;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human prostate specific nucleic acid #103.  
XX  
KW Prostate specific polypeptide; metastasis; prostate cancer; cancer;  
KW non-cancerous prostate disease; gene therapy; gene; ds.  
XX  
OS Homo sapiens.

XX WO200238810-A2.  
XX 16-MAY-2002.  
XX  
XX 06-NOV-2001; 2001WO-US047001.  
XX  
XX 06-NOV-2000; 2000US-0246109P.  
XX  
XX (DIAD-) DIADEXUS INC.  
XX  
XX Sun Y, Recipon H, Chen S, Liu C;  
XX  
XX WPI; 2002-471559/50.  
XX  
XX New prostate specific polypeptides and polynucleotides useful for  
PT detecting, diagnosing, monitoring, treating, staging and predicting  
PT cancers in humans having cancer and non-cancerous prostate disease.  
XX  
XX Claim 1; Page 203-204; 267pp; English.

XX The invention describes an isolated prostate specific polypeptide (I) and  
XX nucleic acid (II) encoding it and are useful for diagnosing and  
XX monitoring the presence and metastases of prostate cancer in a patient.  
XX (I), (II) and an antibody to (II) are useful in quantitative and  
XX qualitative diagnostic assays and methods for detecting, diagnosing,  
XX monitoring, treating, staging and predicting cancer in humans having  
XX cancer or may have the risk of developing cancer. (I) and (II) are also  
XX useful for determining non-cancerous prostate disease, by measuring  
XX their expression levels and/or structural alterations; for determining  
XX the sample that has prostate tissue-like characteristics or is a prostate  
XX tissue; as an element in an array or a multi-analyte test to recognise  
XX expression patterns associated with prostate cancer and other prostate  
XX related disorders; and as elements in a computer program for pattern  
XX recognition of prostate disorders. (I) and (II) are useful for producing

CC engineered prostate tissue for treatment and research. (II) is useful for  
CC producing transgenic animals and cells and also in gene therapy. This  
CC sequence represents a prostate specific nucleic acid described in the  
CC invention

XX Sequence 1450 BP; 341 A; 370 C; 418 G; 321 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 1450;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 619 AAAAAACACACCACTGAGCTGGG 643

RESULT 24  
ABK95452/c  
ID ABK95452 standard; DNA; 2027 BP.

XX  
AC ABK95452;

XX 24-SEP-2002 (first entry)

XX DNA encoding novel G-protein coupled receptor related protein #11.

XX G protein coupled receptor; GPCR; olfactory receptor;  
XX cell signal processing disorder; metabolic pathway modulation;  
XX cardiomyopathy; atherosclerosis; diabetes; developmental disease;  
XX immune disease; taste disorder; scent detectability disorder; obesity;  
XX Burkitt's lymphoma; corticosterogenic disease; infectious disease; pain;  
XX signal transduction pathway disorder; metabolic pathway disorder;  
XX retinal disease; metabolic disorder; cancer; Parkinson's disease;  
XX acute heart failure; urinary retention; osteoporosis; Crohn's disease;  
XX ulcer; allergy; neurological disorder; genetic disorder; transplantation;  
XX fertility; Pancreatitis; Hyperthyroidism; Endometriosis;  
XX forensic biology; transgenic animal; gene; ds.

XX Mus musculus.

XX WO200240539-A2.

XX 23-MAY-2002.

XX 16-OCT-2001; 2001WO-US032256.

XX 16-OCT-2000; 2000US-0240704P.

XX 26-OCT-2000; 2000US-0243497P.

XX 31-OCT-2000; 2000US-024542P.

XX 03-NOV-2000; 2000US-0245484P.

XX 12-DEC-2000; 2000US-0255017P.

XX 17-JAN-2001; 2001US-0262159P.

XX 22-JAN-2001; 2001US-0263216P.

XX 25-JAN-2001; 2001US-0263340P.

XX 12-FEB-2001; 2001US-0264118P.

XX 15-FEB-2001; 2001US-0268225P.

XX 27-JUL-2001; 2001US-0269031P.

XX 27-JUL-2001; 2001US-0308203P.

XX (CURA-) CURAGEN CORP.

XX Kekuda R, Spytek KA, Casman SJ, Zerhusen BD, Li L, Tchernev VT;

XX Colman SD, Ballinger RA, Padigar M, Wolenc AR, Shenoy SG;

XX Edinger SR, Gerlach V, Gangoli EA, Macdougall JR, Smithson G;

XX Peyman JA, Stone DJ, Gunther E, Ellerman K, Grosse WM, Alsobrook JP;

XX Lepley DM, Burgess CB;

XX WPI; 2002-500205/53.

XX P-PSDB; ABG66934.

XX Novel G protein coupled receptor especially olfactory receptor

PT polypeptides and nucleic acids for diagnosing and treating

PT atherosclerosis, cardiomyopathy and diabetes.

XX Claim 8; Page 43; 309pp; English.

PS The invention describes an isolated G protein coupled receptor X (GPCR1-

XX 12) polypeptide, especially an olfactory receptor. GPCR polypeptides are

CC useful for identifying an agent that binds to the polypeptide and for

CC identifying a candidate substance or ligand molecules interacting with an

CC olfactory receptor polypeptide. The polypeptide, (I) and (II) are also

CC useful for treating diseases and disorders related to cell signal

CC processing and metabolic pathway modulation e.g. cardiomyopathy,

CC atherosclerosis and diabetes, and developmental diseases, immune

CC corticosteroid disease, signal transduction pathway disorders,

CC metabolic pathway disorders, retinal diseases, metabolic disorders,

CC obesity, infectious disease, pain, cancer, Parkinson's disease, acute

CC heart failure, urinary retention, osteoporosis, Crohn's disease, ulcers,

CC allergies, neurological disorders, genetic disorders, transplantation,

CC fertility, Pancreatitis, Hyperthyroidism and Endometriosis. GPCR

CC sequences are also useful for identifying a cell or tissue type in a

CC biological sample, to amplify DNA sequences from very small biological

CC samples such as tissues e.g. hair or skin or body fluids in forensic

CC biology. Cells comprising (I) are useful for producing non-human

CC transgenic animals for studying the function and/or activity of GPCR

CC protein and for identifying and/or evaluating modulators of GPCR protein

CC activity. This sequence encodes a novel G-protein coupled receptor

CC described in the invention

XX Sequence 2027 BP; 480 A; 548 C; 505 G; 494 T; 0 U; 0 Other;

SQ

Query Match 74.4%; Score 18.6; DB 6; Length 2027;

Best Local Similarity 84.0%; Pred. No. 3e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25

||||| ||||||| |||

Db 1215 AAAAAACACCACCTGAGCTGGG 1191

RESULT 25

ABK40203/c

ID ABK40203 standard; cDNA; 2028 BP.

XX

AC ABK40203;

XX

DT 21-MAY-2002 (first entry)

XX

DE Human G protein-coupled receptor (GPCR) 9 cDNA #1.

XX

KW Human; G protein-coupled receptor; GPCR; cell signalling processing; ss;

KW cardiomyopathy; atherosclerosis; developmental disease; immune disease;

KW non insulin-dependent diabetes mellitus; NIDDM1; bacterial infection;

KW fungal infection; protozoal infection; viral infection; adenocarcinoma;

KW human immunodeficiency virus; HIV; cancer; lymphoma; anorexia; asthma;

KW neurodegenerative disorder; Parkinson's disease; acute heart failure;

KW hypotension; hypertension; osteoporosis; Crohn's disease; allergy; ulcer;

KW multiple sclerosis; osteodystrophy; angina pectoris; schizophrenia; gene;

KW myocardial infarction; neurological disorder; anxiety; manic depression;

KW delirium; dementia; Huntington's disease; Alzheimer's disease; obesity;

KW Tourette's syndrome; metabolic disorder.

XX

OS Homo sapiens.

XX WO200202637-A2.

PN

XX

PD 10-JAN-2002.

XX

PF 02-JUL-2001; 2001WO-US021174.

XX

PR 30-JUN-2000; 2000US-0215483P.

PR 30-JUN-2000; 2000US-0215495P.

PR 06-JUL-2000; 2000US-0216117P.

PR 25-JUL-2000; 2000US-0220589P.

PR 02-NOV-2000; 2000US-0245294P.

PR

PR 10-JAN-2001; 2001US-0260729P.

PR 10-JAN-2001; 2001US-0260851P.

PR 26-FEB-2001; 2001US-0271673P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Spaderna SK, Padigaru M, Spytek KA, Casman S, Rastelli L;

PI Mayankar U, Tchernev V;

XX

DR WPI; 2002-164524/21.

DR P-PSDB; AAU85873.

XX

PT Novel polypeptides and nucleic acids of G protein coupled receptor useful

PT for diagnosing and treating diseases e.g. anorexia, asthma, Parkinson's

PT disease, acute heart failure and infections.

XX

PS Claim 9; Page 56-57; 202pp; English.

XX

CC The invention relates to human G protein coupled receptor (GPCR)

CC polypeptides and the polynucleotides encoding them. GPCR polypeptides are

CC useful for treating or preventing a GPCR-associated disorder which is

CC related to cell signalling processing and metabolic pathway modulation,

CC which includes cardiomyopathy and atherosclerosis in a human. The

CC sequences are useful in treatment of patients suffering from

CC developmental diseases, immune diseases, non insulin-dependent diabetes

CC mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections

CC (e.g. infections caused by HIV), cancer (e.g. adenocarcinoma, lymphoma),

CC anorexia, asthma, neurodegenerative disorders (e.g. Parkinson's disease),

CC acute heart failure, hypotension, hypertension, osteoporosis, Crohn's

CC disease, multiple sclerosis, osteodystrophy, angina pectoris, myocardial

CC infarction, ulcers, allergies, neurological disorders including anxiety,

CC schizophrenia, manic depression, delirium, dementia, Huntington's

CC disease, Alzheimer's disease, Tourette's syndrome, metabolic disorders

CC and obesity. This sequence represents cDNA encoding a GPCR polypeptide of

CC the invention

XX

SQ Sequence 2028 BP; 480 A; 549 C; 505 G; 494 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 2028;

Best Local Similarity 84.0%; Pred. No. 3e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25

||||| ||||||| |||

Db 1216 AAAAAACACCACCTGAGCTGGG 1192

RESULT 26

ABZ58229/c

ID ABZ58229 standard; cDNA; 2312 BP.

XX

AC ABZ58229;

XX

DT 28-APR-2003 (first entry)

XX

DE Human G-protein coupled receptor GCRC-6 (olfactory receptor) cDNA.

XX

KW GCRC-6; G-protein coupled receptor; olfactory receptor; receptor; human;

KW vasotropic; cardiant; cardiovascular; antiarteriosclerotic; antianginal;

KW antiinflammatory; cerebroprotective; antiemetic; antidiabetic; antigout;

KW neuroprotective; nootropic; tranquilizer; osteopathic; cytostatic;

KW hypotensive; gastrointestinal; antiulcer; antianaemic; antithyroid;

KW anticonvulsant; antiparkinsonian; anorectic; virucide; olfaction; taste;

KW gene therapy; gene; ss.

XX

OS Homo sapiens.

XX

XX

FX Key Location/Qualifiers

FT CDS 1371..2288

FT /\*tag= a

FT /product= "Human GCRC-6"

XX

PN WO2003000859-A2.

```
XX 03-JAN-2003.
PD
XX
PF 20-JUN-2002; 2002WO-US020036.
XX
XX 22-JUN-2001; 2001US-0300494P.
PR 13-JUL-2001; 2001US-0305326P.
PR 27-JUL-2001; 2001US-0308165P.
PR 03-AUG-2001; 2001US-0310115P.
PR 10-AUG-2001; 2001US-0311413P.
PR 24-AUG-2001; 2001US-03114679P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Baughn MR, Walia NK, Richardson TW, Ison CH, Yang J, Emerling BM;
PI Thornton M, Borowsky ML, Ramkumar J, Yue H, Graul RC, Hafalia AJA,
PI Lu DAM, Lu Y, Kallick DA, Gandhi AR, Warren BA, Lee EA;
PI Swarnakar A, Walsh RT, Yao MG, Griffin JA, Au-Young JK, Khan FA;
PI Lee SY, Elliott VS;
XX
XX WPI; 2003-221414/21.
DR P-PSDB; ABP72212.
XX
XX New human G-protein coupled receptor proteins and polynucleotides, useful
PT for diagnosing, treating or preventing e.g. ischemic heart disease,
PT Crohn's disease, AIDS, stroke, Creutzfeldt-Jakob disease, leukemia or
PT lymphoma.
XX
XX Claim 5; Page 171-172; 182pp; English.
XX
XX The present sequence is that of a cDNA clone, denoted Incyte
CC Polynucleotide ID No. 55153201CB1, encoding novel human G-protein coupled
CC receptor GCRC-6. The encoded protein is an olfactory receptor, being 73%
CC identical from residue 1 to 296 to a murine olfactory receptor, and
CC including a 7-transmembrane receptor (rhodopsin family) domain. The
CC invention provides GCRC-1 to -26 polynucleotides (see AB258224-49),
CC polypeptides (see ABP72207-32) and antibodies. These are useful for the
CC diagnosis, treatment or prevention of neural disorders of olfaction, cell
CC proliferative, neurological, cardiovascular, gastrointestinal,
CC autoimmune, inflammatory, and metabolic disorders, and viral infections.
CC They are also used in claimed methods of screening for agonist and
CC antagonist compounds, in treating a disease or condition associated with
CC decreased expression or overexpression of a functional GCRC, in
CC assessing the toxicity of a test compound, in the construction of a
CC microarray and generation of an expression profile, and in the
CC identification of a compound which modulates, mimics and/or blocks an
CC olfactory and/or taste sensation
XX
XX Sequence 2312 BP; 564 A; 636 C; 568 G; 544 T; 0 U; 0 Other;
SQ
Query Match 74.4%; Score 18.6; DB 8; Length 2312;
Best Local Similarity 84.0%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACACACCACTGACCTGGG 25
Db 2115 AAAAAACACCACTGCGGTG 2091
RESULT 27
AAK68477
ID AAK68477 standard; DNA; 2464 BP.
XX
XX AAK68477;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23289.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
```

XX  
PN  
XX  
PD  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.



```
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249277P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
XX
PA
XX
```

```
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides;
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 23289; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK4703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 2464 BP; 529 A; 658 C; 791 G; 486 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 2464;
Best Local Similarity 84.0%; Pred. NO. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
DB 1751 AAAAAACCAACAAACTGAAGCTGGG 1775

RESULT 28
AAK68479
ID AAK68479 standard; DNA; 2464 BP.
AC AAK68479;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:23291.
XX
KW Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 17-JUL-2000; 2000US-0217487P.
```

PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225211P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225269P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246611P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246613P.
PR	14-AUG-2000;	2000US-0225759P.	PR	17-NOV-2000;	2000US-0249207P.
PR	18-AUG-2000;	2000US-0226279P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249209P.
PR	22-AUG-2000;	2000US-0226688P.	PR	17-NOV-2000;	2000US-0249210P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249211P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249212P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249215P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249216P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249217P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249218P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249244P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249245P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0231414P.	PR	01-DEC-2000;	2000US-0250160P.
PR	08-SEP-2000;	2000US-0232080P.	PR	01-DEC-2000;	2000US-0250391P.
PR	08-SEP-2000;	2000US-0232081P.	PR	05-DEC-2000;	2000US-0251030P.
PR	13-SEP-2000;	2000US-0231968P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232398P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232399P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251990P.
PR	14-SEP-2000;	2000US-0233065P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234223P.	PR	05-JAN-2001;	2001US-0

CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 2464 BP; 529 A; 661 C; 790 G; 484 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 2464;  
Best Local Similarity 84.0%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

Db 1751 AAAAAACACCAAACTGACTGGG 1775

RESULT 29

AAAL04040/c

ID AAAL04040 standard; DNA; 2464 BP.

XX AAAL04040;

AC XX

DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 6728.

DE Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.

XX Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184666P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234224P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.

```
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Baraah SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 6728; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 2464 BP; 486 A; 791 C; 658 G; 529 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 4; Length 2464;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AAAAAACAACCACTGAGCTGGG 25
Db 714 AAAAAACAACCACTGAGCTGGG 690
|||||||
|||||||
RESULT 30
AAL04041/c
ID AAL04041 standard; DNA; 2464 BP.
XX
AC AAL04041;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6729.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 23-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 12-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
XX 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
XX 29-SEP-2000; 2000US-0236327P.
```





XX AC ADI16310;  
XX DT 22-APR-2004 (first entry)  
XX DE Human nucleic acid-associated protein (NAAP) coding sequence #45.  
XX KW human; nucleic acid-associated protein; NAAP; autoimmune disorder;  
XX KW inflammatory disorder; AIDS; allergy; infection; metabolic disorder;  
XX KW obesity; reproductive disorder; infertility; neurological disorder;  
XX KW Parkinson's disease; Alzheimer's disease; cardiovascular disorder;  
XX KW myocardial infarction; hypertension; eye disorder;  
XX KW cell proliferative disease; cancer; ds; gene.  
XX OS Homo sapiens.  
XX PN WO2003094848-A2.  
XX PD 20-NOV-2003.  
XX PF 09-MAY-2003; 2003WO-US014450.  
XX PR 10-MAY-2002; 2002US-0379843P.  
XX PR 24-MAY-2002; 2002US-0383457P.  
XX PR 31-MAY-2002; 2002US-0384699P.  
XX PR 06-JUN-2002; 2002US-0387265P.  
XX PA (INCY-) INCYTE CORP.  
XX PI Kable AE, Elliott VS, Tran UK, Ramkumar J, Marquis JP, Chawla NK;  
PI Richardson TW, Bulloch SA, Khare R, Lee SY, Lal PG, Tang YT, Yue H;  
PI Swarnakar A, Becha SD, Hafalia AUA, Chang H, Baughn MR, Borowsky ML;  
PI Gietzen KJ, He A, Forsythe IU, Sprague WW, Blake JU, Warren BA;  
PI Mason PM, Ison CH, Lindquist EA, Wilson AD, Jin P;  
XX WPI; 2004-011999/01.  
XX DR P-PSDB; ADP16260.  
XX PT New human nucleic acid associated proteins and polynucleotides, useful  
XX PT for diagnosing, preventing or treating diseases or conditions associated  
XX PT with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or  
XX PT stroke.  
XX PS Claim 5; SEQ ID NO 95; 400pp; English.  
XX SS The invention comprises the amino acid and coding sequences of human  
XX CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
XX CC the invention are useful in diagnosing, preventing and treating  
XX CC diseases/conditions associated with altered expression of NAAP, such as:  
XX CC autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections  
XX CC (e.g. bacterial and viral), metabolic disorders (e.g. obesity),  
XX CC reproductive disorders (e.g. infertility), neurological disorders (e.g.  
XX CC Parkinson's disease and Alzheimer's disease), cardiovascular disorders  
XX CC (e.g. myocardial infarction and hypertension), eye disorders, or cell  
XX CC proliferative diseases (e.g. cancer). The present DNA sequence encodes a  
XX CC human NAAP protein of the invention.  
XX SQ Sequence 3030 BP; 865 A; 705 C; 787 G; 673 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 12; Length 3030;  
Best Local Similarity 84.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 1 AAAAAACACCCACTGAGCTGGG 25  
Db 2988 AAAAAACACCCACTGAGCTGGG 3012  
RESULT 35  
ADR25447  
ID ADR25447 standard; DNA; 3154 BP.  
XX ADR25447;  
AC

XX DT 21-OCT-2004 (first entry)  
XX DE Breast cancer prognosis marker #1308.  
XX KW ds; breast cancer; prognosis; gene expression; diagnosis.  
XX OS Homo sapiens.  
XX PN WO2004065545-A2.  
XX PD 05-AUG-2004.  
XX PF 15-JAN-2004; 2004WO-US001100.  
XX PR 15-JAN-2003; 2003US-00342887.  
XX PA (ROSE-) ROSETTA INPHARMATICS LLC.  
XX PA (NECA-) NETHERLANDS CANCER INST.  
XX PI Van't Veer LJ, He Y;  
XX WPI; 2004-593473/57.  
XX DR Classifying a breast cancer patient according to prognosis comprises  
XX PT determining the similarity between the level of expression of each of  
XX PT five genes in a cell sample taken from patient, to control levels.  
XX PS Disclosure; SEQ ID NO 1308; 226pp; English.  
XX CC The invention relates to a method of classifying a breast cancer patient  
XX CC according to prognosis by determining the similarity between the level of  
XX CC expression of each of five genes for which markers are listed in the  
XX CC specification, in a cell sample taken from the breast cancer patient, to  
XX CC control levels of expression for each respective five genes to obtain a  
XX CC patient similarity value. The methods are useful for classifying a breast  
XX CC cancer patient according to prognosis. Kits and computer program products  
XX CC are useful for data analysis using the diagnostic, prognostic and  
XX CC statistical methods of the invention. This sequence corresponds to a  
XX CC marker used in the method of the invention.  
XX SQ Sequence 3154 BP; 901 A; 734 C; 811 G; 708 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 13; Length 3154;  
Best Local Similarity 84.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 1 AAAAAACACCCACTGAGCTGGG 25  
Db 3112 AAAAAACACCCACTGAGCTGGG 3136  
RESULT 36  
AAZ57865  
ID AAZ57865 standard; cDNA; 3190 BP.  
XX AC AAZ57865;  
XX DT 11-APR-2000 (first entry)  
XX DE Protein regulating gene expression PRGE-27 cDNA clone 1399169.  
XX KW Protein regulating gene expression; PRGE-27; human; cell proliferation;  
XX KW antiproliferative; inflammation; antinflammatory; gene therapy;  
XX KW diagnosis; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT CDS 14..2869  
XX FT /\*tag= a  
XX PN WO9964596-A2.



```
XX 16-DEC-1999.
PD
XX
PF 11-JUN-1999; 99WO-US013281.
XX
XX 12-JUN-1998; 98US-0089029P.
PR
XX 29-JUL-1998; 98US-0094575P.
PR
XX 14-OCT-1998; 98US-0104624P.
PR
XX
XX (INCY-) INCYTE PHARM INC.
PA
XX
XX Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;
PI Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DAM;
XX
XX WPI; 2000-116543/10.
DR P-PSDB; AAY58634.
XX
XX New human polypeptides that regulate gene expression, for treatment,
PT prevention and diagnosis of, e.g. cancer.
PT
XX
XX Claim 9; Page 144-145; 150pp; English.
PS
XX
XX The present sequence is that of Incyte clone 1399169 encoding new human
CC protein regulating gene expression PRGE-27 (see AAY58634). The cDNA was
CC initially isolated from brain tumour cDNA library BRAITUT08, and the full
CC length sequence assembled from overlapping clones from a number of
CC libraries. PRGE-27 is expressed in haematopoietic/immune, reproductive
CC and gastrointestinal tissues associated with cell proliferative and
CC inflammation diseases, disorders or conditions. It is characterized as a
CC bromodomain protein. The invention provides PRGE polypeptides (see
CC AAY58608-38) and polynucleotides (see AAZ57839-69), expression vectors,
CC host cells, antibodies, agonists and antagonists. It also provides
CC methods for diagnosing, treating or preventing disorders associated with
CC expression of PRGE. Polynucleotides are also used as sources of probes
CC and primers for diagnosis and monitoring of disease, also for detecting
CC related sequences and in gene mapping
XX
XX Sequence 3190 BP; 918 A; 739 C; 822 G; 711 T; 0 U; 0 Other;
SQ
Query Match 74.4%; Score 18.6; DB 3; Length 3190;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACACACCACCTGAGCTGGG 25
Db 3138 AAAAAACACACCACCTGAGTGAATGTG 3162
RESULT 37
ADX05939
ID ADX05939 standard; DNA; 3226 BP.
XX
XX
AC ADX05939;
XX
XX 21-APR-2005 (first entry)
DT
XX
XX Cyclin-dependent kinase modulation biomarker DNA SEQ ID NO 504.
DE
XX
XX cytostatic; cyclin-dependent kinase; cdk; biomarker; gene; ds.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2005012875-A2.
FN
XX
XX 10-FEB-2005.
PD
XX
XX 29-JUL-2004; 2004WO-US024424.
PF
XX
XX 29-JUL-2003; 2003US-0490890P.
PR
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA
XX
XX Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;
PI
```

```
XX WPI; 2005-163068/17.
DR P-PSDB; ADX05940.
XX
XX Biomarkers useful for predicting or determining the response of a mammal
PT to a cancer treatment comprising administration of a modulator of cyclin-
PT dependent kinase activity.
XX
XX Claim 5; SEQ ID NO 504; 141pp; English.
PS
XX
XX This invention describes a novel method of predicting or determining
CC whether a mammal will respond or is responding to an anti-cancer agent
CC that modulates cyclin-dependent kinase (cdk) activity. The method
CC comprises measuring the level of one or more biomarkers selected from
CC 2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC invention is utilized in a kit for determining or predicting whether
CC patient would be susceptible or resistant to treatment by an agent
CC modulating cdk activity. The invention also describes a method for
CC utilizing individualized genetic profiles for treating diseases and
CC disorders based on patient's response and molecular level, specialized
CC microarrays comprising the biomarkers described, antibodies directed
CC against the biomarkers and a cell culture model to identify biomarkers.
CC The cdk modulator is preferably N-5-[(5-(1,1-Dimethylethyl)-2-
CC oxazolyl)methylthio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC tartaric acid salt. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. This
CC sequence encodes a biomarker used in the method of the invention.
XX
XX Sequence 3226 BP; 940 A; 743 C; 828 G; 715 T; 0 U; 0 Other;
SQ
Query Match 74.4%; Score 18.6; DB 14; Length 3226;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACACACCACCTGAGCTGGG 25
Db 3153 AAAAAACACACCACCTGAGTGAATGTG 3177
RESULT 38
ACN42607
ID ACN42607 standard; cDNA; 3512 BP.
XX
XX ACN42607;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human diagnostic and therapeutic polynucleotide SEQ ID NO:1482.
DE
XX
XX ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KW dithp.
XX
XX Homo sapiens.
OS
XX
XX WO2004023973-A2.
FN
XX
XX 25-MAR-2004.
PD
XX
XX 12-SEP-2003; 2003WO-US028227.
PF
XX
XX 12-SEP-2002; 2002US-0410259P.
PR
XX
XX 12-SEP-2002; 2002US-0410260P.
PR
XX
XX (INCY-) INCYTE CORP.
PA
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic WM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LJ;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI
```

PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;  
PI Patury S, Shi X, Suarez C;  
XX WPI; 2004-329368/30.  
DR P-PSDB; A8W83955.  
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful  
PT in diagnosing a condition, disease or disorder associated with human  
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or  
XX in gene mapping.  
PS Claim 1; Page; 190pp; English.  
XX The invention relates to novel diagnostic and therapeutic polynucleotides  
CC selected from one of the 2722 sequences defined in the specification. A  
CC polynucleotide of the invention may have a use in gene therapy. The human  
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be  
CC used to diagnose a particular condition, disease or disorder associated  
CC with human molecules, e.g. cell proliferative disorders,  
CC autoimmune/inflammatory disorder, developmental disorder, endocrine  
CC disorder, neurological disorders, gastrointestinal disorders, or  
CC infections caused by virus, bacteria, fungi or parasite. The dithp  
CC molecules may also be used in genetic mapping, in identifying individuals  
CC from minute biological samples, in detecting single nucleotide  
CC polymorphisms, as molecular weight markers, and for somatic or germline  
CC gene therapy. The present sequence represents a dithp polynucleotide of  
CC the invention. Note: the sequence data for this patent is not represented  
CC in the printed specification, but was obtained in electronic format  
CC directly from WIPO at [www.wipo.int/pct/en/sequences/listing.htm](http://www.wipo.int/pct/en/sequences/listing.htm)  
XX  
SQ Sequence 3512 BP; 998 A; 816 C; 899 G; 799 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 13; Length 3512;  
Best Local Similarity 84.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACACCACTGAGCTGGG 25  
DB 3454 AAAAAACACACCACTGAAATGTG 3478  
RESULT 39  
ADR07844  
ID ADR07844 standard; cDNA; 4281 BP.  
AC ADR07844;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Full length human cDNA useful for treating neurological disease Seq 1350.  
XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;  
KW osteoporosis; neurological disease; Alzheimer's disease;  
KW Parkinson's disease; dementia; short memory; cancer;  
KW sense or motor function; emotional reaction; fear response; panic;  
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;  
KW tranquiliser.  
XX  
OS Homo sapiens.  
XX  
PN EP1447413-A2.  
XX  
PD 18-AUG-2004.  
XX  
PF 12-FEB-2004; 2004EP-00003145.  
XX  
XX 14-FEB-2003; 2003JP-00102207.  
PR 09-MAY-2003; 2003JP-00131452.  
XX  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;  
PI Wakamatsu A, Ishii S, Nagai K, Irie R;

XX WPI; 2004-583265/57.  
DR P-PSDB; ADR09800.  
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.  
XX  
PS Claim 1; SEQ ID NO 1350; 2686pp; English.  
XX This invention relates to novel, isolated full length human cDNA  
CC molecules and the encoded proteins thereof. Specifically, it refers to  
CC cDNA clones obtained by an oligo-capping method, where none of these  
CC clones are identical to any known human mRNAs. The present invention  
CC describes an immunoassay to identify agonists and antagonists, as well as  
CC antibodies, antisense molecules and siRNAs that can all be used to bind  
CC to and modulate expression of the cDNA molecules. As such, these  
CC molecules are useful for diagnostic markers or therapeutic targets for  
CC the various diseases or morbid states. In particular, they are useful in  
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
CC disease, Parkinson's disease, dementia, short memory and various cancers,  
CC as well as for maintaining equilibrium of sense or motor function, and  
CC for treating emotional reaction, fear response and panic. Accordingly,  
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
CC cyostatic and tranquilliser activities. This polynucleotide is a full  
CC length human cDNA sequence of the invention. NOTE: This sequence is not  
CC given in the sequence listing of the specification but can be obtained on  
CC CD-ROM from the European Patent Office, Vienna Sub-office.  
XX  
SQ Sequence 4281 BP; 812 A; 1421 C; 1143 G; 905 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 13; Length 4281;  
Best Local Similarity 84.0%; Pred. No. 3.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACACCACTGAGCTGGG 25  
DB 2459 AAGAAAGAACCTCCCTGAGCTGGG 2483  
RESULT 40  
AAK67429/c  
ID AAK67429 standard; DNA; 9840 BP.  
XX  
AC AAK67429;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22241.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.





one of the polynucleotides cited above. The polypeptides and polynucleotides of the invention have antiinflammatory, cytostatic, and antimicrobial activities. The novel polynucleotide may be used to treat disorders by gene therapy. The polypeptides and polynucleotides are useful for treating inflammation, leukaemias, nervous system disorders, or infections. This polynucleotide sequence represents contiguous DNA derived from one of the 235 novel isolated polynucleotides of the invention.

Sequence 16578 BP; 4635 A; 3859 C; 3976 G; 4108 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 13; Length 16578;  
Best Local Similarity 84.0%; Pred. No. 4.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCCACCACTGAGCTGGG 25  
|||||  
DB 16262 AAAAAACAACCCACCACTGCGGTG 16238  
|||||

RESULT 43  
ABZ74034/c  
ID ABZ74034 standard; DNA; 39119 BP.  
XX AC  
XX ABZ74034;  
DT 12-MAY-2003 (first entry)  
XX AC  
XX  
DE Secreted protein gene 163 genomic fragment HJPC08, SEQ ID NO:1181.  
XX  
KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
KW drug screening; chromosome identification; chromosome mapping;  
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;  
KW antianaemic; vulnerary; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200277013-A2.  
PN  
XX  
XX  
PD 03-OCT-2002.  
XX  
XX 26-MAR-2002; 2002WO-US009370.  
XX  
XX 27-MAR-2001; 2001US-0278650P.  
PR 12-SEP-2001; 2001US-00950082.  
PR 12-SEP-2001; 2001US-00950083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX WPI; 2003-040578/03.  
DR  
XX  
XX New human secreted proteins and nucleic acids, useful for detecting or treating cancer or other hyperproliferative disorders, autoimmune disorders, inflammatory disorders, HIV disease, hepatitis or anemia.  
PT  
XX  
XX Disclosure; Page 1836-1845; 2474pp; English.  
PS  
XX  
XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted protein genes, and ABP00947-ABP01363 represent the proteins they encode.  
CC  
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The invention also encompasses antibodies specific for the secreted proteins, the use of the secreted proteins in drug screening and recombinant vectors and host cells comprising a nucleic acid of the invention. The secreted proteins are thought to be involved in biological activities associated with cellular signalling, cellular differentiation, cell migration, prohormone activation and neurotransmitter activity. The secreted proteins, nucleic acids encoding them, antibodies or antibody fragments specific for the secreted proteins, and modulators of protein activity are useful for diagnosing or treating cancers or other

CC	hyperproliferative disorders. Additionally, the secreted proteins and
CC	their nucleic acids may also be used in the treatment of autoimmune
CC	disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC	(acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC	wound healing. Nucleic acids of the invention may be used for chromosome
CC	identification, chromosome mapping, in gene therapy, for identifying
CC	individuals from minute biological samples, as hybridisation probes, and
CC	as molecular weight markers. The present sequence represents a human
CC	secreted protein genomic fragment referred to in the disclosure of the
CC	invention
XX	
XX	Sequence 39119 BP; 11690 A; 8925 C; 7906 G; 10598 T; 0 U; 0 Other;
XX	
XX	Query Match 74.4%; Score 18.6; DB 8; Length 39119;
XX	Best Local Similarity 84.0%; Pred. NO. 4.7e+02;
XX	Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX	
Qy	1 AAAAAACACCCACTGAGCTGGG 25
Db	17118 AAAAAACACCCACTGAGTAAATGTG 17094
RESULT 44	
ADA98641/c	
ID ADA98641 standard; DNA; 39119 BP.	
XX AC ADA98641;	
XX AC	
DT 20-NOV-2003 (first entry)	
XX	
DE Human secreted protein-related DNA sequence #234.	
XX	
KW human; secreted protein; cardiovascular disorder; arrhythmia;	
KW atherosclerosis; stroke; endocarditis; congestive heart failure;	
KW rheumatic heart disease; cardiomyopathy; hemorrhoids; varicose veins;	
KW migraine; thrombosis; neural disorder; immune system disorder;	
KW muscular disorder; reproductive disorder; gastrointestinal disorder;	
KW pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.	
XX	
OS Homo sapiens.	
XX	
FN WO2003004623-A2.	
XX	
PD 16-JAN-2003.	
XX	
PF 26-MAR-2002; 2002WO-US009922.	
XX	
PR 27-MAR-2001; 2001US-0278650P.	
PR 12-SEP-2001; 2001US-00950082.	
PR 12-SEP-2001; 2001US-00950083.	
XX	
PA (HUMA-) HUMAN GENOME SCI INC.	
XX	
PI Rosen CA, Ruben SM;	
XX	
DR WPI; 2003-247946/24.	
XX	
PT New human secreted polypeptide and nucleic acid molecules, useful for	
PT diagnosing, preventing, prognosticating or treating cardiovascular	
PT disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or	
PT thrombosis).	
XX	
PS Disclosure; SEQ ID NO 750; 1572pp; English.	
XX	
CC The invention comprises the amino acid and coding sequence of human	
CC secreted proteins. The DNA and protein sequences of the invention are	
CC useful in the treatment of cardiovascular disorders, such as: arrhythmia,	
CC atherosclerosis, stroke, endocarditis, congestive heart failure,	
CC rheumatic heart disease, cardiomyopathy, hemorrhoids, varicose veins,	
CC migraine, or thrombosis. The DNA and protein sequences may also be used	
CC for treating or preventing: neural disorders, immune system disorders,	
CC muscular disorders, reproductive disorders, gastrointestinal disorders,	
CC pulmonary disorders, renal disorders, proliferative disorders and/or	

CC cancerous diseases. The present DNA sequence is used in the  
CC exemplification of the invention. NOTE: The present sequence is shown on  
CC the WIPO website.

XX SQ Sequence 39119 BP; 11690 A; 8925 C; 7906 G; 10598 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 8; Length 39119;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACACACCACCTGAGCTGGG 25  
Db 17118 AAAAAACACACCACCTGAGCTGGG 17094

RESULT 45  
ADC20764/c  
ID ADC20764 standard; DNA; 39119 BP.  
XX AC ADC20764;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human secreted protein-related DNA sequence #182.  
XX KW gene therapy; human; secreted protein; haemopoietic disorder;  
KW hematological disorder; anemia; haemophilia; inflammatory disorder;  
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;  
KW leukaemia; wound healing; epithelial cell proliferation disorder;  
KW immune disorder; autoimmune disorder; asthmatic disorder;  
KW cardiovascular disorder; atherosclerosis; myocarditis;  
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;  
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200292787-A2.  
XX PD 21-NOV-2002.  
XX PF 26-MAR-2002; 2002WO-US009257.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX PT WPI; 2003-129287/12.  
XX DR New human secreted proteins and nucleic acid molecules, useful for  
PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
PT preventing or treating hematopoietic or hematologic disorders, e.g.  
PT anemia or hemophilia.  
XX PS Disclosure; SEQ ID NO 718; 1512pp; English.  
XX CC The invention comprises the amino acid and coding sequences of human  
CC secreted proteins. The DNA and protein sequences of the invention are  
CC useful for detecting, preventing, diagnosing, prognosticating, treating  
CC or ameliorating; haematopoietic or haematological disorders (e.g. anaemia  
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease  
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);  
CC wound healing and disorders of epithelial cell proliferation; immune  
CC disorders (e.g. autoimmune disorders and asthmatic disorders);  
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);  
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);  
CC and gastrointestinal disorders (e.g. duodenal ulcers and  
CC gastroenteritis). The present DNA sequence was used in the  
CC exemplification of the invention.

XX SQ Sequence 39119 BP; 11690 A; 8925 C; 7906 G; 10598 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 10; Length 39119;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACACACCACCTGAGCTGGG 25  
Db 17118 AAAAAACACACCACCTGAGCTGGG 17094

RESULT 46  
ABZ67621/c  
ID ABZ67621 standard; DNA; 39119 BP.  
XX AC ABZ67621;  
XX DT 26-MAR-2003 (first entry)  
XX DE Human secreted protein encoding genomic DNA SEQ ID NO 1144.  
XX KW Human; secreted protein; neutropenic; neuroprotective; cytostatic;  
KW viricide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;  
KW vulnary; antibacterial; antiparkinsonian; antiskilling; antianemic;  
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;  
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;  
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;  
KW cardiovascular disorder; neurological disease; nephrotropic;  
KW gene therapy; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200277186-A2.  
XX PD 03-OCT-2002.  
XX PF 26-MAR-2002; 2002WO-US009188.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX PT WPI; 2003-040583/03.  
XX DR New human secreted proteins encoded by genes contained in cDNA clones  
PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,  
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or  
PT West Nile fever.  
XX PS Disclosure; Page 1804-1814; 2423pp; English.  
XX CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the  
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,  
CC treating or ameliorating medical conditions e.g. by protein or gene  
CC therapy. The genes are isolated from a range of human tissues disclosed  
CC in the specification. The nucleic acids, proteins, antibodies and  
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections  
XX SQ Sequence 39119 BP; 11690 A; 8925 C; 7906 G; 10598 T; 0 U; 0 Other;



```
Query Match      74.4%; Score 18.6; DB 10; Length 39119;
Best Local Similarity 84.0%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACAACACACTGAGCTGG 25
Db 17118 AAAAAACAACACACTGGAATGTG 17094

RESULT 47
ABS98185/c
ID ABS98185 standard; DNA; 53099 BP.
XX
AC ABS98185;
XX
DT 23-DEC-2002 (first entry)
XX
DE Human multidrug resistance gene exons 4-28 sequence.
XX
XX Human; db; cytochrome P450 A1; CYP4501A1; UGT2B4; MDR1;
KW cytochrome P450 A2; CYP4501A2; cytochrome P450 02E; CYP45002E1; LTF;
KW adrenergic receptor beta1; ADRB1; aryl hydrocarbon; AHR; MRP3; NR1I2;
KW aryl hydrocarbon receptor nuclear translocator; ARNT; cathepsin S; CTSS;
KW cyclooxygenase 2; COX2; diazepam binding inhibitor; DBI; haematological;
KW epoxide hydroxylase 2; EPHX2; 5-lipoxygenase activating protein; FLAP;
KW glutathione-S-transferase 12; GST12; histamine-N-methyl transferase;
KW HMT; kallikrein 2; KLK2; nicotinamide-N-methyl transferase; NNMT;
KW NADPH quinone oxidoreductase 2; NQO2; sulfoltransferase thermolabile; STM;
KW UDP-glucuronosyl transferase 2B4; UDP-glucuronosyl transferase 2B7;
KW UGT2B7; UDP-glucuronosyl transferase; UGT2B15; urokinase receptor; uPA;
KW multidrug resistance 1; lactotransferrin; orphan nuclear receptor;
KW multidrug resistance associated protein 3; cancer; prostate;
KW acetylcholine muscarinic receptor; CHMR1; CHMR2; CHMR3; CHMR4; CHMR5;
KW altered drug metabolism; cardiovascular function; colorectal tumour;
KW central nervous system; pulmonary; immunological; SNP;
KW single nucleotide polymorphism.
XX
XX Homo sapiens.
XX
XX WO200257410-A2.
XX
XX 25-JUL-2002.
XX
XX 28-NOV-2001; 2001WO-US044838.
XX
XX 28-NOV-2000; 2000US-00724389.
XX
XX (DNAS-) DNA SCI LAB INC.
XX
XX Guida M, Hall J;
XX
XX WPI; 2002-698522/75.
XX
XX Isolated nucleic acid molecules having polymorphisms in known human genes
XX e.g. cytochrome P450 and cathepsin S useful as genetic linkage markers
XX for locating, identifying and characterizing the genes responsible for
XX disorder-related traits.
XX
XX Example 22; Page 437-467; 714pp; English.
XX
XX This invention relates to the sequence of an isolated nucleic acid
XX molecule comprising at least one base variation from that of a known
XX human cytochrome P450 A1 (CYP4501A1), cytochrome P450 A2 (CYP4501A2),
XX cytochrome P450 02E1 (CYP45002E1), adrenergic receptor beta1 (ADRB1),
XX aryl hydrocarbon (AHR), aryl hydrocarbon receptor nuclear translocator
XX (ARNT), cathepsin S (CTSS), cyclooxygenase 2 (COX2), diazepam binding
XX inhibitor (DBI), epoxide hydroxylase 2 (EPHX2), 5-lipoxygenase activating
XX protein (FLAP), glutathione-S-transferase 12 (GST12), histamine-N-methyl
XX transferase (NNMT), (kallikrein 2) KLK2, nicotinamide -N-methyl
XX sulfoltransferase thermolabile (STM), UDP-glucuronosyl transferase 2B4
XX (UGT2B4), UDP-glucuronosyl transferase 2B7 (UGT2B7), UDP-glucuronosyl
XX transferase (UGT2B15), urokinase receptor (uPA), multidrug resistance 1
```



FT	variation	4523	/tag= g	/standard_name= "Single nucleotide polymorphism (SNP)"	FT	/tag= ae	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	4645	/tag= h		FT	18538	
FT	variation	4646	/tag= i		FT	/tag= af	
FT	variation	5830	/tag= j		FT	18810	
FT	variation	6556	/tag= k		FT	/tag= ag	
FT	variation	6568	/tag= l		FT	20282	
FT	variation	6570	/tag= m		FT	/tag= ah	
FT	variation	9351	/tag= n		FT	20495	
FT	variation	10874	/tag= o		FT	/tag= ai	
FT	variation	10926	/tag= p		FT	20633	
FT	variation	11462	/tag= q		FT	/tag= aj	
FT	variation	12415	/tag= r		FT	20659	
FT	variation	12668	/tag= s		FT	/tag= ak	
FT	variation	13886	/tag= t		FT	21603	
FT	variation	15005	/tag= u		FT	/tag= al	
FT	variation	15387	/tag= v		FT	/tag= am	
FT	variation	15539	/tag= w		FT	24075	
FT	variation	15761	/tag= x		FT	/tag= an	
FT	variation	16969	/tag= y		FT	/tag= ao	
FT	variation	17124	/tag= z		FT	/tag= ap	
FT	variation	17369	/tag= aa		FT	25145	
FT	variation	17523	/tag= ab		FT	/tag= aq	
FT	variation	18158	/tag= ac		FT	25222	
FT	variation	18220	/tag= ad		FT	/tag= ar	
FT	variation	18224	/tag= ae		FT	25288	
FT	variation				FT	/tag= as	
FT	variation				FT	26080	
FT	variation				FT	/tag= at	
FT	variation				FT	26556	
FT	variation				FT	/tag= au	
FT	variation				FT	27416	
FT	variation				FT	/tag= av	
FT	variation				FT	27432	
FT	variation				FT	/tag= aw	
FT	variation				FT	29841	
FT	variation				FT	/tag= ax	
FT	variation				FT	29955	
FT	variation				FT	/tag= ay	
FT	variation				FT	30160	
FT	variation				FT	/tag= az	
FT	variation				FT	30294	
FT	variation				FT	/tag= ba	
FT	variation				FT	31039	
FT	variation				FT	/tag= bb	
FT	variation				FT	32071	
FT	variation				FT	/tag= bc	



Qy 1 AAAAACAACCACTGAGCTGG 25  
Db 121849 AAAAGACACCCCTCACTGAGCTGG 121873

Search completed: February 3, 2006, 21:56:24  
Job time : 212.111 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-3

Perfect score: 25

Sequence: 1 aaaaacaccaccactgagctggg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_est3:\*

4: gb\_est4:\*

5: gb\_est5:\*

6: gb\_est6:\*

7: gb\_est7:\*

8: gb\_est8:\*

9: gb\_est9:\*

10: gb\_est10:\*

11: gb\_est11:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	403	1	AI551075
C 2	25	100.0	412	5	BY627047
C 3	25	100.0	444	1	AW909560
C 4	25	100.0	499	6	CB249608
C 5	25	100.0	517	3	BM935380
C 6	25	100.0	664	1	BB035891
C 7	25	100.0	1993	4	AK077834
C 8	25	100.0	3000	4	AK086675
C 9	23.4	93.6	430	5	BY635162
C 10	23.4	93.6	450	2	BF464082
C 11	23.4	93.6	478	2	BB760039
C 12	22	88.0	395	5	BY650318
C 13	21	84.0	394	5	BY674654
C 14	20.2	80.8	172	7	CV525379
C 15	20.2	80.8	256	7	CV525125
C 16	20.2	80.8	375	5	BY646584
C 17	20.2	80.8	577	1	AL633505
C 18	20.2	80.8	596	5	EX746867
C 19	20.2	80.8	600	5	BQ526123
C 20	20.2	80.8	606	1	AL681607
C 21	20.2	80.8	653	1	AL860221
C 22	20.2	80.8	654	1	AL854589

CR584874	CR584874	CR584874	654	7	CR584874
AL885471	AL885471	AL885471	656	1	AL885471
CR359039	JGI_XZT41	CR359039	719	8	CR359039
CR408645	JGI_XZT31	CR408645	744	8	CR408645
BI978107	pf10_Old	BI978107	758	3	BI978107
CR504564	JGI_XZG41	CR504564	761	8	CR504564
CF219540	AGENCOURT	CF219540	767	6	CF219540
DN099260	JGI_CABE8	DN099260	768	8	DN099260
CR833540	JGI_CAA12	CR833540	770	8	CR833540
CR416784	JGI_XZG64	CR416784	773	8	CR416784
DR834532	JGI_CABC5	DR834532	781	8	DR834532
BY729979	BY729979	BY729979	788	5	BY729979
DN100573	JGI_CABE9	DN100573	807	5	DN100573
CR852874	JGI_CAA19	CR852874	818	8	CR852874
CF218992	AGENCOURT	CF218992	820	6	CF218992
BY758175	BY758175	BY758175	822	5	BY758175
CR403435	CR403435	CR403435	822	8	CR403435
CR444938	CR444938	CR444938	830	7	CR444938
DR863945	JGI_CABG7	DR863945	836	8	DR863945
BY782362	BY782362	BY782362	844	5	BY782362
CR423927	CR423927	CR423927	848	7	CR423927
CR441514	CR441514	CR441514	851	7	CR441514
BY771898	BY771898	BY771898	862	5	BY771898
CR568362	CR568362	CR568362	862	5	CR568362
BY755834	BY755834	BY755834	864	5	BY755834
BY750203	BY750203	BY750203	869	5	BY750203
CR568361	CR568361	CR568361	872	7	CR568361
BY749420	BY749420	BY749420	877	5	BY749420
BY763188	BY763188	BY763188	948	5	BY763188
BM461492	AGENCOURT	BM461492	1582	3	BM461492
CN228322	RUB039B11	CN228322	421	7	CN228322
BU211361	604159668	BU211361	454	5	BU211361
CN384034	4117981_B	CN384034	483	7	CN384034
BU339642	603515146	BU339642	573	5	BU339642
BU422511	603954064	BU422511	603	5	BU422511
BU335085	603500361	BU335085	629	5	BU335085
DN929637	4155309_B	DN929637	642	8	DN929637
BU287148	604162713	BU287148	671	5	BU287148
EX276390	EX276390	EX276390	691	5	EX276390
BU310187	603536987	BU310187	700	5	BU310187
BU382139	603856506	BU382139	702	5	BU382139
EX276389	EX276389	EX276389	723	5	EX276389
BU120866	603145783	BU120866	727	5	BU120866
BU378670	603589836	BU378670	738	5	BU378670
BU404481	604137919	BU404481	762	5	BU404481
BU336237	603869978	BU336237	779	5	BU336237
CF288358	AGENCOURT	CF288358	793	6	CF288358
BU116519	603138472	BU116519	866	5	BU116519
CB561500	AGENCOURT	CB561500	921	6	CB561500
BU128033	603112909	BU128033	935	5	BU128033
CC225679	CH261-87C	CC225679	948	9	CC225679
AG400048	Mus muscu	AG400048	721	10	AG400048
AZ753304	RPCI-24-8	AZ753304	760	9	AZ753304
T50566	YD76B11.r1	T50566	227	8	T50566
EX476195	JGI_XZG19	EX476195	339	8	EX476195
CE125562	tigr-g88-	CE125562	489	9	CE125562
B2113085	CH230-233	B2113085	513	9	B2113085
BP282422	BP282422	BP282422	581	3	BP282422
AQ922900	RPCI-23-2	AQ922900	602	9	AQ922900
AL652889	AL652889	AL652889	654	1	AL652889
BZ252589	CH230-402	BZ252589	685	9	BZ252589
CR973038	CR973038	CR973038	716	11	CR973038
BX140525	Danio rer	BX140525	792	10	BX140525
BZ147085	CH230-368	BZ147085	793	9	BZ147085
CR814198	GR0AAA39A	CR814198	832	11	CR814198
AJ927585	StrPus37	AJ927585	977	1	AJ927585
CD322861	StrPus37	CD322861	518	6	CD322861
EX143389	1283032_N	EX143389	600	8	EX143389
CC757729	ZMMBB014	CC757729	369	9	CC757729
CC175325	ZMMBB029	CC175325	369	9	CC175325
CZ306000	ZMMBF0001	CZ306000	448	10	CZ306000

```
c 96 18.8 75.2 485 9 A0845099 an46a08 J
c 97 18.8 75.2 486 9 C983799 ZUAC988TH
c 98 18.8 75.2 495 6 CB713124 AGNNUC.N
c 99 18.8 75.2 501 9 A0845100 an46a09 J
c 100 18.8 75.2 510 9 A0844270 an29c08 J
c 101 18.8 75.2 519 9 A0845079 an44g06 J
c 102 18.8 75.2 530 6 CB719046 AGNNUC.N
c 103 18.8 75.2 543 10 C2341486 ZMBBF0111
c 104 18.8 75.2 559 9 BZ776093 ih96d01.g
c 105 18.8 75.2 577 9 BH182062 020_O_19-
c 106 18.8 75.2 577 11 CNS07NJM AL161912 T3 end of
c 107 18.8 75.2 581 10 CG024655 ZMBB0056
c 108 18.8 75.2 586 6 CB077287 hJ52a03.g
c 109 18.8 75.2 624 10 C0793416 ZMBB0030
c 110 18.8 75.2 632 10 C0713690 ZMBB0000
c 111 18.8 75.2 645 6 CA351434 622504 NC
c 112 18.8 75.2 645 8 CG029229 1341104 N
c 113 18.8 75.2 653 9 AZ416349 1M0191K05
c 114 18.8 75.2 656 10 CL355109 RPCI44.40
c 115 18.8 75.2 676 10 C0736482 ZMBB0032
c 116 18.8 75.2 692 10 AG126291 Pan trogl
c 117 18.8 75.2 698 10 CG010669 ZUA8V47TV
c 118 18.8 75.2 713 9 CC977408 ZUA8725TV
c 119 18.8 75.2 713 11 CR867166 Sus scrof
c 120 18.8 75.2 715 6 CA368573 644796 NC
c 121 18.8 75.2 725 9 CC461247 ZMBB0036
c 122 18.8 75.2 727 9 BZ651474 OGBAD90TM
c 123 18.8 75.2 733 9 BZ988638 PUCAW06TD
c 124 18.8 75.2 737 10 CG795584 ZMBB0032
c 125 18.8 75.2 745 10 CW004239 ZMBB0000
c 126 18.8 75.2 752 6 CD102779 AGENCOURT
c 127 18.8 75.2 766 10 C2274186 ZMBBF0055
c 128 18.8 75.2 770 10 CG872643 ZMBB0027
c 129 18.8 75.2 775 9 CC175852 ZMBB0029
c 130 18.8 75.2 782 11 CR867346 Sus scrof
c 131 18.8 75.2 792 10 C2350903 ZMBBF0086
c 132 18.8 75.2 798 9 CC753941 ZMBB0013
c 133 18.8 75.2 807 10 C2304352 ZMBBF0082
c 134 18.8 75.2 809 7 CK316740 SB02019A2
c 135 18.8 75.2 813 9 CC703882 OGUET43TH
c 136 18.8 75.2 819 10 C2388932 ZMBBF0165
c 137 18.8 75.2 828 10 C2338950 ZMBBF0107
c 138 18.8 75.2 837 9 CC333520 CG1BE01TH
c 139 18.8 75.2 845 9 CC974619 ZUA8E63TV
c 140 18.8 75.2 848 10 C2296777 ZMBBF0066
c 141 18.8 75.2 849 9 CC825858 ZMBB0017
c 142 18.8 75.2 849 9 CC834256 ZMBB0018
c 143 18.8 75.2 853 10 C2276273 ZMBBF0059
c 144 18.8 75.2 859 10 CL281555 ZMBB0062
c 145 18.8 75.2 875 9 CC992763 ZUA8E23TV
c 146 18.8 75.2 876 10 CW002921 ZMBB0000
c 147 18.8 75.2 876 10 CC333526 CG1BE01TV
c 148 18.8 75.2 879 9 CC333526 CG223383 OGLAD06TH
c 149 18.8 75.2 884 10 CG223383 OGLAD06TH
c 150 18.8 75.2 886 9 CC978249 ZUA8H79TV
```

## ALIGNMENTS

```
RESULT 1
A1551075/c 403 bp mRNA linear EST 23-MAR-1999
LOCUS IMAGE:1277237 3', mRNA sequence.
DEFINITION vx35h03.x1 Stratagene mouse lung 937302 Mus musculus cDNA clone
ACCESSION A1551075
VERSION A1551075.1 GI:4403438
KEYWORDS Mus musculus (house mouse)
SOURCE EST.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Murinae; Mus.
```

REFERENCE  
AUTHORS

1 (bases 1 to 403)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:669037

This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
Possible reversed clone: polyT not found  
High quality sequence stop: 388.

FEATURES  
source

1..403  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1277237"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse lung 937302"  
/note="Organ: lung; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCAGTCTTTTCTTTTCTTTT 3'"

## ORIGIN

Query Match 100.0%; Score 25; DB 1; Length 403;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAACACCACTGAGCTGGG 25  
|||||  
Db 292 AAAAAAACACCACTGAGCTGGG 268  
|||||

## RESULT 2

BY627047 412 bp mRNA linear EST 15-DEC-2002  
BY627047 RIKEN full-length enriched, visual cortex Mus musculus  
cDNA clone K430014A06 3', mRNA sequence.

ACCESSION BY627047  
VERSION BY627047.1 GI:26962229  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 412)  
AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kavasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perteaux,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Iehli,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

**TITLE** Analysis of the mouse transcriptome based on functional annotation

**JOURNAL** Nature 420, 563-573 (2002)

**PUBMED** 12466851

**COMMENT** Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Iehli,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Michela Fagiolini and Takao K. Henschi (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hiroawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
source

Location/Qualifiers  
1. .412  
/organism="Mus musculus"  
/mol\_type="mrna"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K430014A06"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

**ORIGIN**

Query Match 100.0%; Score 25; DB 5; Length 412;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY**

**Db**

**RESULT 3**

**LOCUS** AW909560/c

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**

1 AAAAAAACACCACTGAGCTGGG 25

|||||

14 AAAAAAACACCACTGAGCTGGG 38

AW909560 444 bp mRNA linear EST 25-MAY-2000  
ur76h03.x1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3156245 3',  
mrna sequence.

AW909560 GI:8074797

AW909560.1

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 444)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other ESTs: ur76h03.y1

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

MG1:1059001

Possible reversed clone: polyT not found

Seq primer: -40UP from Gibco

High quality sequence stop: 407.

Location/Qualifiers

source

1. .444  
/organism="Mus musculus"  
/mol\_type="mrna"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3156245"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/clone\_lib="NCI CGAP Mam6"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

**ORIGIN**

Query Match 100.0%; Score 25; DB 1; Length 444;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY**

**Db**

**RESULT 4**

**LOCUS** CB249608

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

CB249608 499 bp mRNA linear EST 15-JUL-2003  
UI-M-EX0-by1-g-22-0-UI.r1 NIH\_EMAP\_EX0 Mus musculus cDNA clone  
IMAGE:5719725 5', mRNA sequence.

CB249608

CB249608.1 GI:28389208

EST.

```

SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
REFERENCE        Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         Eukaryota; Muroidea; Muridae; Murinae; Mus.
JOURNAL         Sciurognathi; Muroidea; Muridae; Murinae; Mus.
TITLE           1 (bases 1 to 499)
COMMENT         NIH-MGC http://mgc.nci.nih.gov/
                National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgsbbs@mail.nih.gov
                Tissue Procurement: Dr. James Lin, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                This clone was contributed by the Brain Molecular Anatomy Project
                (BMAP)
                The following repetitive elements were found in this cDNA
                sequence: 3-108, >B1-FHSINE/Alu 122-195, >ID3#SINE/ID 333-363,
                >AT rich#Low_complexity 374-499, >B1_MM#SINE/Alu (matched
                complement)
                Seq primer: pyX-5.
                Location/Qualifiers
                1..499
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:5719725"
                /tissue_type="whole brain"
                /dev_stage="embryo 15.5 dpc"
                /lab_host="DH10B (T1 phage resistant)"
                /clone_lib="NIH BMAP EX0"
                /note="Organ: brain; Vector: pyX-Asc; Site 1: EcoR I;
                Site 2: Not I; The library was constructed according to
                Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                1996. Denatured mRNA was size fractionated on a 1% agarose
                gel. First strand cDNA synthesis was primed with an
                oligo-dT primer containing a Not I site. Double stranded
                cDNA was size selected according to mRNA size fraction.
                ligated with EcoR I adaptor, digested with Not I, and then
                cloned directionally into pyX-Asc vector. The library tag
                sequence located between the Not I site and the polyA
                tail, is GTCGCTGGAA. This library was created for the
                University of Iowa Mouse Brain Molecular Anatomy Project
                (BMAP). 'Gene Discovery in the Developing Mouse Nervous
                System', supported by National Institutes of Mental Health
                (NIMH), Hemin Chin, Ph.D., program coordinator."

FEATURES        source
source          1..499
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:5719725"
                /tissue_type="whole brain"
                /dev_stage="embryo 15.5 dpc"
                /lab_host="DH10B (T1 phage resistant)"
                /clone_lib="NIH BMAP EX0"
                /note="Organ: brain; Vector: pyX-Asc; Site 1: EcoR I;
                Site 2: Not I; The library was constructed according to
                Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                1996. Denatured mRNA was size fractionated on a 1% agarose
                gel. First strand cDNA synthesis was primed with an
                oligo-dT primer containing a Not I site. Double stranded
                cDNA was size selected according to mRNA size fraction.
                ligated with EcoR I adaptor, digested with Not I, and then
                cloned directionally into pyX-Asc vector. The library tag
                sequence located between the Not I site and the polyA
                tail, is GTCGCTGGAA. This library was created for the
                University of Iowa Mouse Brain Molecular Anatomy Project
                (BMAP). 'Gene Discovery in the Developing Mouse Nervous
                System', supported by National Institutes of Mental Health
                (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      100.0%; Score 25; DB 6; Length 499;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
    |||||
Db 105 AAAAAACACCACTGAGCTGGG 129
    |||||

RESULT 5
BM935380/c
LOCUS           517 bp mRNA linear EST 29-APR-2002
DEFINITION     UI-M-CGOp-bnw-a-11-0-UI.r1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
                UI-M-CGOp-bnw-a-11-0-UI 5', mRNA sequence.
ACCESSION      BM935380
VERSION        BM935380.1 GI:19394532
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
Query Match      100.0%; Score 25; DB 3; Length 517;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
    |||||
Db 131 AAAAAACACCACTGAGCTGGG 107
    |||||

RESULT 6
BB035891
LOCUS           664 bp mRNA linear EST 19-OCT-2001
DEFINITION     BB035891 RIKEN full-length enriched, 13 days embryo forelimb Mus
                musculus cDNA clone 5930406J05 3', mRNA sequence.
ACCESSION      BB035891
VERSION        BB035891.2 GI:16258342
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.
                1 (bases 1 to 664)
                Arahawa, T., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T.,
                Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,
                Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

```

```

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 517)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
889548
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of
Medicine
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).
The following repetitive elements were found in this cDNA
sequence: 41-114, >ID3#SINE/ID (matched complement) 128-283,
>B1_MM#SINE/Alu (matched complement) 296-367, >PB1D7#SINE/Alu
(matched complement)
Seq primer: M13 REVERSE.
                Location/Qualifiers
                1..517
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UI-M-CGOp-bnw-a-11-0-UI"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="NIH BMAP Ret4 S2"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site 1: Not I; Site 2: Eco RI; The
                NIH_BMAP Ret4_S2 library is a subtracted library,
                ultimately derived from mouse retina tissue libraries at
                various stages of development. For a detailed description
                of the library from which this clone was derived, please
                visit our web site at brainest.eng.uiowa.edu. The tissue
                for this library was contributed by Dr. Xin-Yuan Fu, Yale
                University School of Medicine"

FEATURES        source
source          1..517
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UI-M-CGOp-bnw-a-11-0-UI"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="NIH BMAP Ret4 S2"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site 1: Not I; Site 2: Eco RI; The
                NIH_BMAP Ret4_S2 library is a subtracted library,
                ultimately derived from mouse retina tissue libraries at
                various stages of development. For a detailed description
                of the library from which this clone was derived, please
                visit our web site at brainest.eng.uiowa.edu. The tissue
                for this library was contributed by Dr. Xin-Yuan Fu, Yale
                University School of Medicine"

ORIGIN
Query Match      100.0%; Score 25; DB 3; Length 517;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
    |||||
Db 131 AAAAAACACCACTGAGCTGGG 107
    |||||

RESULT 6
BB035891
LOCUS           664 bp mRNA linear EST 19-OCT-2001
DEFINITION     BB035891 RIKEN full-length enriched, 13 days embryo forelimb Mus
                musculus cDNA clone 5930406J05 3', mRNA sequence.
ACCESSION      BB035891
VERSION        BB035891.2 GI:16258342
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.
                1 (bases 1 to 664)
                Arahawa, T., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T.,
                Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,
                Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

```





Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/.

**FEATURES**  
 source  
 1..1993  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:5930406J05"  
 /db\_xref="taxon:10090"  
 /clone="5930406J05"  
 /tissue\_type="forelimb"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="13 days embryo"  
 1..1993  
 /note="unknown EST (GB|BB035891, evidence: BLASTN, 99%, match=663)"

**ORIGIN**  
 Query Match . 100.0%; Score 25; DB 4; Length 1993;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25  
 |||||||||  
 Db 1595 AAAAAACACCACTGAGCTGGG 1619  
 |||||||||

**RESULT 8**  
 AK086675 3000 bp mRNA linear HTC 03-APR-2004  
 LOCUS Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:D930045C14 product:unknown EST, full insert sequence.  
 DEFINITION  
 ACCESSION AK086675  
 VERSION AK086675.1 GI:26103681  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE**  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 10349636

**REFERENCE**  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 11042159

**REFERENCE**  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 11076861

**REFERENCE**  
 4  
 AUTHORS THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

**REFERENCE**  
 5  
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 3000)

**REFERENCE**  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saio, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

**COMMENT**  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/.

**FEATURES**  
 source  
 1..3000  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:D930045C14"  
 /db\_xref="taxon:10090"  
 /clone="D930045C14"  
 /tissue\_type="head"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="15 days embryo"  
 1..3000  
 /note="unknown EST (GB|BB035891, evidence: BLASTN, 99%, match=663)"

**misc\_feature**  
 1..3000  
 /note="unknown EST (GB|BB035891, evidence: BLASTN, 99%, match=663)"

**ORIGIN**  
 Query Match 100.0%; Score 25; DB 4; Length 3000;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||||||  
 Db 857 AAAAAACCACTGAGCTGGG 881  
 |||||||||

```

RESULT 9
BY635162
LOCUS
DEFINITION
  BY635162 RIKEN full-length enriched, visual cortex Mus musculus
  cDNA clone K430066P18 3', mRNA sequence.
ACCESSION
  BY635162
VERSION
  BY635162.1 GI:26970344
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Mus musculus
  Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
  Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
  Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I.,
  Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A.,
  Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C.,
  Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H.,
  Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V.,
  Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A.,
  Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T.,
  Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S.,
  Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A.,
  Kawai H., Kawasawa Y., Kedzierzki R.M., King B.L., Konagaya A.,
  Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R.,
  Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T.,
  Numata K., Okido T., Favan W.J., Pertea G., Pesole G.,
  Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
  Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
  Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
  Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
  Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y.,
  Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I.,
  Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P.,
  Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M.,
  Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K.,
  Arakawa T., Fukuda S., Hara A., Haehizume W., Imotani K., Ishii Y.,
  Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K.,
  Shingawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S.,
  Rogers J., Birney B. and Hayashizaki Y.
  Analysis of the mouse transcriptome based on functional annotation
  of 60,770 full-length cDNAs
  Nature 420, 563-573 (2002)
JOURNAL
  1246851
PUBMED
  Contact: Yoshihide Hayashizaki
  Laboratory for Genome Exploration Research Group, RIKEN Genomic
  Sciences Center (GSC), Yokohama Institute
  The Institute of Physical and Chemical Research (RIKEN)
  1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
  Tel: 81-45-503-9222
  Fax: 81-45-503-9216
  Email: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/
  Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
  Hirozane T., Imotani K., Ishii Y., Itoh M., Kawai J., Konno H.,
  Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
  Ono M., Sakai K., Sakazume N., Sasaki D., Sato K., Shibata K.,
  Shiraki T., Tagami M., Waki K., Watahiki A., Muramatsu M. and
  Hayashizaki Y. Direct Submission
  Computational Analysis of Full-Length Mouse cDNAs Compared with
  Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new
  genes. Genome Res. 10 (10), 1617-1630 (2000)
  RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer. Genome Res.
  10 (11), 1757-1771 (2000)
  Computer-based methods for the mouse full-length cDNA
  encyclopedia: real-time sequence clustering for construction of a
  nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
  cDNA library was prepared and sequenced in Mouse Genome
  Encyclopedia Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in RIKEN.

```

Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Fagiolini and Takao K. Hensch ( Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan ) whose  
assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

Location/Qualifiers

1. .430

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="K430066P18"

/tissue\_type="visual cortex"

/clone\_lib="RIKEN full-length enriched, visual cortex"

FEATURES

source

Query Match 93.6%; Score 23.4; DB 5; Length 430;

Best Local Similarity 96.0%; Pred. No. 15;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACTGAGCTGGG 25

31 AAGAAACACCCACTGAGCTGGG 55

Db

RESULT 10

BF464082

LOCUS

DEFINITION

UI-M-CGOp-bnw-a-11-0-UI.s1 NIH BMAP Ret4 S2 Mus musculus cDNA clone

UI-M-CGOp-bnw-a-11-0-UI 3', mRNA sequence.

ACCESSION

BF464082

VERSION

BF464082.1 GI:11533265

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 450)

Ronaldo M.F., Lennon G. and Soares M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mst@mail.nih.gov

Oligo-dt track not found, Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements

The following repetitive elements were found in this cDNA sequence:

159-230, >PBLD7#SINE/Alu 243-398, >B1\_MM#SINE/Alu

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1. .450

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-CGOp-bnw-a-11-0-UI"

/lab\_host="DH10B (Life Technologies)"

FEATURES

source

/clone lib="NIH BMAP Ret4 S2"  
 /note="Vector: pTT3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NIH\_BMAP\_Ret4\_S2 library is a subtracted library,  
 ultimately derived from mouse retina tissue libraries at  
 various stages of development. For a detailed description  
 of the library from which this clone was derived, please  
 visit our web site at brainest.eng.uiowa.edu. The tissue  
 for this library was contributed by Dr. Xin-Yuan Fu, Yale  
 University School of Medicine  
 TAG\_SEQ=None found"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 450;  
 Best Local Similarity 96.0%; Pred. No. 16;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACAACCACTGAGCTGGG 25

Db 395 ACANAACAACCACTGAGCTGGG 419

## RESULT 11

BB760039

LOCUS BB760039 478 bp mRNA linear EST 17-OCT-2001  
 DEFINITION BB760039 RIKEN full-length enriched, melanocyte Mus musculus cDNA  
 clone G270124E08 3', mRNA sequence.

ACCESSION BB760039

VERSION BB760039.1 GI:16206348

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Mus.

1 (bases 1 to 478)

## REFERENCE

AUTHORS

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
 Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,  
 Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
 Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,  
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,  
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Tanaka, T., Tomaru, A., Toyota, T., Watahiki, A., Yasunishi, A.,  
 Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
 2001)

## TITLE

Unpublished (2001)

## JOURNAL

## COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
 Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,  
 and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. .  
 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.

## FEATURES

source

Location/Qualifiers  
 1..478  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G270124E08"  
 /cell\_type="melanocyte"  
 /clone\_lib="RIKEN full-length enriched, melanocyte"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 478;

Best Local Similarity 96.0%; Pred. No. 16;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACAACCACTGAGCTGGG 25

Db 79 AAGAAACAACCACTGAGCTGGG 103

## RESULT 12

BY650318

LOCUS BY650318 395 bp mRNA linear EST 16-DEC-2002  
 DEFINITION BY650318 RIKEN full-length enriched, visual cortex Mus musculus  
 cDNA clone K530047N21 3', mRNA sequence.

ACCESSION BY650318

VERSION BY650318.1 GI:27007288

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Mus.

1 (bases 1 to 395)

## REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osaoto, N., Saito, R., Suzuki, H., Yamanaka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,  
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beissel, K. W., Blake, J. A., Bradt, D., Brusic, V.,  
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,  
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,  
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,  
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,  
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Sempile, C. A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
 Vekardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L. G., Wyszynski, B., Yang, Y., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
 Rogers, J., Birney, E., and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

## COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirotsawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1..395  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K530047N21"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

## ORIGIN

Query Match 88.0%; Score 22; DB 5; Length 395;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACACACCACCTGAGCTGGG 25

|||||

1 AACACACCACCTGAGCTGGG 22

## RESULT 13

BY674654

LOCUS

DEFINITION  
Ratke's pouches Mus musculus cDNA clone K820041010 3', mRNA  
sequence.

ACCESSION

BY674654

VERSION

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikiado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusis, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,

Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gunzich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., Mckenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomika, M., Verardo, C., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, K., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

## TITLE

JOURNAL

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1..394  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="K820041010"  
/tissue\_type="Rathke's pouches"  
/dev\_stage="14.5 days embryo df/df"  
/clone\_lib="RIKEN full-length enriched, 14.5 days embryo df/df Rathke's pouches"

## ORIGIN

Query Match

84.0%; Score 21; DB 5; Length 394;

Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AACAAACCACTGAGCTGGG 25  
 |||||  
 Db 1 AACAAACCACTGAGCTGGG 21  
 |||||

RESULT 14  
 CV525379  
 LOCUS  
 DEFINITION  
 Mdlv4011c18.y1 Mdlv Malus x domestica cDNA clone Mdlv4011c18 5',  
 similar to TR:Q9ZNS3 Q9ZNS3 RIBOSOMAL PROTEIN S27. ;, mRNA  
 sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Malus x domestica  
 Malus x domestica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids 1; Rosales; Rosaceae; Maloideae; Malus.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 172)  
 Korbán,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
 Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orvis,K.,  
 Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
 Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,  
 Tsagarisvili,R., Kennedy,S., Waterston,R. and Wilson,R.,  
 Apple Functional Genomics grant - NSF 0321702  
 Unpublished (2004)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Schuyler S. Korban  
 Apple Functional Genomics grant - NSF 0321702  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Library materials provided by: Schuyler S. Korban Library  
 constructed by: D.O.Gonzalez/L. Vodkin Library sequenced by:  
 Washington University Genome Sequencing Center  
 WashU EST name: aan94b09.y1  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 1.  
 Location/Qualifiers

FEATURES  
 source

1..172  
 /organism="Malus x domestica"  
 /mol\_type="mRNA"  
 /cultivar="GoldRush"  
 /db\_xref="taxon:3750"  
 /clone="Mdlv4011c18"  
 /tissue\_type="Leaf"  
 /lab\_host="DH10B ampicillin resistant"  
 /clone\_lib="Mdlv"

/note="Vector: pSPORT 1; Site 1: NotI; Site 2: Sal I;  
 Total RNA was extracted from freeze dried leaf tissue  
 Stage I, using a standard Phenol chloroform extraction  
 method. Poly(A)+mRNA was isolated from total RNA using the  
 PolyAtract mRNA Isolation system III (Promega). The  
 library was prepared using the Invitrogen Life  
 Technologies, Superscript Plasmid System with gateway  
 technology for cDNA Synthesis and cloning. Complementary  
 DNA was synthesized from mRNA using an anchored Poly (dT)  
 sequence with a NotI restriction site. Sal I linker  
 adapters were ligated to the blunt ended cDNA fragments  
 followed by restriction with NotI. The cDNA fragments were  
 directionally cloned into the NotI-SalI restriction site  
 of the pSPORT 1 vector. The ligated cDNA fragments were  
 transformed into E. coli Electromax DH10B host cells.  
 Transformation efficiency: 8.5E+07 Average insert Size by  
 PCR: 1100 bp"

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 172;  
 Best Local Similarity 88.0%; Pred. No. 3.9e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACAACCACTGAGCTGGG 25  
 |||||  
 Db 12 AAAAAACAACCACTGAGCTGGG 36  
 |||||

RESULT 15  
 CV525125  
 LOCUS  
 DEFINITION  
 Mdlv4010g21.y1 Mdlv Malus x domestica cDNA clone Mdlv4010g21 5',  
 similar to TR:Q9ZNS3 Q9ZNS3 RIBOSOMAL PROTEIN S27. ;, mRNA  
 sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Malus x domestica  
 Malus x domestica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids 1; Rosales; Rosaceae; Maloideae; Malus.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 256)  
 Korbán,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
 Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orvis,K.,  
 Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
 Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,  
 Tsagarisvili,R., Kennedy,S., Waterston,R. and Wilson,R.,  
 Apple Functional Genomics grant - NSF 0321702  
 Unpublished (2004)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Schuyler S. Korban  
 Apple Functional Genomics grant - NSF 0321702  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Library materials provided by: Schuyler S. Korban Library  
 constructed by: D.O.Gonzalez/L. Vodkin Library sequenced by:  
 Washington University Genome Sequencing Center  
 WashU EST name: aan89d11.y1  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 1.  
 Location/Qualifiers

FEATURES  
 source

1..256  
 /organism="Malus x domestica"  
 /mol\_type="mRNA"  
 /cultivar="GoldRush"  
 /db\_xref="taxon:3750"  
 /clone="Mdlv4010g21"  
 /tissue\_type="Leaf"  
 /lab\_host="DH10B ampicillin resistant"  
 /clone\_lib="Mdlv"

/note="Vector: pSPORT 1; Site 1: NotI; Site 2: Sal I;  
 Total RNA was extracted from freeze dried leaf tissue  
 Stage I, using a standard Phenol chloroform extraction  
 method. Poly(A)+mRNA was isolated from total RNA using the  
 PolyAtract mRNA Isolation system III (Promega). The  
 library was prepared using the Invitrogen Life  
 Technologies, Superscript Plasmid System with gateway  
 technology for cDNA Synthesis and cloning. Complementary  
 DNA was synthesized from mRNA using an anchored Poly (dT)  
 sequence with a NotI restriction site. Sal I linker  
 adapters were ligated to the blunt ended cDNA fragments  
 followed by restriction with NotI. The cDNA fragments were  
 directionally cloned into the NotI-SalI restriction site  
 of the pSPORT 1 vector. The ligated cDNA fragments were  
 transformed into E. coli Electromax DH10B host cells.  
 Transformation efficiency: 8.5E+07 Average insert Size by







Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TGas066n23.q1kt7  
Sequencing primer: T7  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13  
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI  
Host: Escherichia coli XLI-blue.

## FEATURES

source

1..596  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TGas066n23"  
/dev\_stage="gastrula (stages 10.5-12 mixed)"  
/lab\_host="Escherichia coli XLI-blue"  
/clone\_lib="XGC-gastrula"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dt primed from sug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 596;  
Best Local Similarity 88.0%; Pred. No. 4.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 260 AAAAAACCACTGACTGGG 284

## RESULT 19

BQ526123/c

LOCUS

DEFINITION BQ526123 600 bp mRNA linear EST 10-JUN-2002  
IMAGE:5381083 5', mRNA sequence.

ACCESSION BQ526123

VERSION BQ526123.1

KEYWORDS GI:21384992

SOURCE EST.

ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 600)

REFERENCE NIH-XGC <http://image.llnl.gov/image/html/xenopuslib.info.shtml>.  
National Institute of Child Health and Human Development, National  
Cancer Institute, Xenopus Gene Collection  
Unpublished (2002)

AUTHORS Contact: Robert Strausberg, Ph.D.

TITLE Email: cgapbs-r@mail.nih.gov

JOURNAL CDNA Library Preparation:

COMMENT DNA Sequencing by: The I.M.A.G.E. Consortium/LLNL

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[info@image.llnl.gov](http://image.llnl.gov)

Plate: LLAM11970 row: I column: 20

Seq primer: M13RP1 reverse primer (ABI).

1..600

Location/Qualifiers

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="IMAGE:5381083"

/tissue\_type="cadpole"

/dev\_stage="embryo, stages 40-45"

/lab\_host="DH10B (phage-resistant)"

/clone lib="NICHD XGC Emb8"  
/note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;  
Cloned unidirectionally. Primer: Oligo dt. Average insert  
size 2.1 kb. Constructed by Invitrogen. Note: This is a  
Xenopus Gene Collection (XGC) library."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 600;  
Best Local Similarity 88.0%; Pred. No. 4.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 258 AAAAAACCACTGTACTGGG 234

## RESULT 20

AL681607/c

LOCUS

DEFINITION AL681607 606 bp mRNA linear EST 10-NOV-2003  
mRNA sequence.

ACCESSION AL681607

VERSION AL681607.2

KEYWORDS GI:38253507

SOURCE EST.

ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 606)

REFERENCE Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

AUTHORS Sanger Xenopus tropicalis EST project 2001 (11\_2003)

TITLE Unpublished (2003)

JOURNAL On Mar 18, 2002 this sequence version replaced gi:19537981.

COMMENT Contact: Taylor R

Sanger Institute

Hinnton, Cambridgeshire, CB10 1SA, UK

Email: [trop@sanger.ac.uk](mailto:trop@sanger.ac.uk)

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13  
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

Host: Escherichia coli XLI-blue

Sanger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: TGas066n23.plcSP6

Sequencing primer: SP6.

Location/Qualifiers

1..606

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TGas066n23"

/dev\_stage="gastrula (stages 10.5-12 mixed)"

/lab\_host="Escherichia coli XLI-blue"

/clone\_lib="XGC-gastrula"

/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dt primed from sug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 606;  
Best Local Similarity 88.0%; Pred. No. 4.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 356 AAAAAACCACTGTACTGGG 332

```
RESULT 21
AL860221/c
LOCUS
DEFINITION
AL860221 XGC-egg Xenopus tropicalis cDNA clone TEGG071011 5', mRNA
sequence.
ACCESSION
AL860221
VERSION
AL860221.2 GI:38655576
KEYWORDS
SOURCE
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 653)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
UNPUBLISHED (2003)
COMMENT
On Sep 15, 2002 this sequence version replaced gi:22880402.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: TEGG071011.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..653
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TEGG071011"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-egg"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
FEATURES
source
ORIGIN
Query Match 80.8%; Score 20.2; DB 1; Length 653;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACACCACCTGAGCTGGG 25
||||| ||||| ||||| ||||| |||||
DB 169 AAAAAACACCACCTGAGCTGGG 145
||||| ||||| ||||| ||||| |||||
RESULT 22
AL854589/c
LOCUS
DEFINITION
AL854589 XGC-egg Xenopus tropicalis cDNA clone TEGG020p23 5', mRNA
sequence.
ACCESSION
AL854589
VERSION
AL854589.2 GI:38630096
KEYWORDS
SOURCE
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 654)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
UNPUBLISHED (2003)
COMMENT
On Sep 15, 2002 this sequence version replaced gi:22874809.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: TEGG020p23.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..654
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TEGG020p23"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-egg"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
FEATURES
source
ORIGIN
Query Match 80.8%; Score 20.2; DB 1; Length 654;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACACCACCTGAGCTGGG 25
||||| ||||| ||||| ||||| |||||
DB 359 AAAAAACACCACCTGAGCTGGG 335
||||| ||||| ||||| ||||| |||||
RESULT 23
CR584874/c
LOCUS
DEFINITION
CR584874 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA045d02
5', mRNA sequence.
ACCESSION
CR584874
VERSION
CR584874.1 GI:50584874
KEYWORDS
SOURCE
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 654)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
UNPUBLISHED (2004)
COMMENT
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: THDA045d02.plkSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
1..654
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA045d02"
/lab_host="Escherichia coli DH10B."
FEATURES
source
```

/clone.lib="XGC-tailbud-head"  
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
 was oligo dt primed from Sug of poly A+ RNA from tailbud  
 head. EcoRI-NotI cut cDNA was then ligated into pCS107  
 with EcoRI at the 5' end and NotI at the 3' end."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 654;  
 Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||

Db 97 AAAAAACACCACTGACTGGG 73  
 |||||

## RESULT 24

AL885471/c  
 LOCUS AL885471 XGC-egg Xenopus tropicalis cDNA clone TEGG04013 5', mRNA  
 DEFINITION  
 VERSION AL885471 656 bp linear EST 03-DEC-2003  
 ACCESSION  
 KEYWORDS  
 SOURCE EST  
 ORGANISM Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 656)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Sep 16, 2002 this sequence version replaced gi:22926745.

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: TEGG04013.p1kSP6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the

5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue.

Location/Qualifiers

## FEATURES

source

1..656

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TEGG04013"

/dev\_stages="egg"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-egg"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from Sug of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end"

## ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 656;  
 Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||

Db 518 AAAAAACACCACTGACTGGG 494  
 |||||

## RESULT 25

CX359039/c

LOCUS

DEFINITION

IMAGE:7580481 5', mRNA sequence.

ACCESSION

VERSION CX359039

KEYWORDS

SOURCE

ORGANISM

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 719)

Richardson, P., Lucas, S., Rokhe, D., Dettler, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute

Unpublished (2004)

Other ESTs: JGI XZT410.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley; http://tropicalis.berkeley.edu/home

cDNA library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LINL;

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone ID and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Plate: XZT 0005 row: c column: 7

High quality sequence stop: 719.

Location/Qualifiers

1..719

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="IMAGE:7580481"

/tissue\_type="whole embryo"

/dev\_stages="Tadpole (st. 36-41)"

/lab\_host="E. coli XL1-Blue derivative, Stratagene

Electrofen-Blue"

/clone\_lib="NIH XGC tropTad5"

/note="Vector: pCS108; Site 1: Sall; Site 2: NotI; Tadpole

library constructed by Russell B. Fletcher in R. Harland's

lab using poly A RNA and oligo dt primers (Invitrogen

SuperScript Plasmid System for cDNA Synthesis and

cloning). Sall (5' end) -NotI (3' end) cDNA was inserted

into vector pCS108

(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

"

## ORIGIN

Query Match

Best Local Similarity

Matches

22; Conservative

0; Mismatches

3; Indels

0; Gaps

0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||

Db 558 AAAAAACACCACTGTACTGGG 534  
 |||||

## RESULT 26

CX408645/c

LOCUS

DEFINITION

IMAGE:7607944 5', mRNA sequence.

ACCESSION

CX408645

CX359039  
 JGI\_XZT410.fwd NIH XGC tropTad5 Xenopus tropicalis cDNA clone  
 IMAGE:7580481 5', mRNA sequence.

CX359039  
 CX359039.1 GI:57127598  
 EST.

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 719)

Richardson, P., Lucas, S., Rokhe, D., Dettler, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute

Unpublished (2004)

Other ESTs: JGI XZT410.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley; http://tropicalis.berkeley.edu/home

cDNA library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LINL;

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone ID and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Plate: XZT 0005 row: c column: 7

High quality sequence stop: 719.

Location/Qualifiers

1..719

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="IMAGE:7580481"

/tissue\_type="whole embryo"

/dev\_stages="Tadpole (st. 36-41)"

/lab\_host="E. coli XL1-Blue derivative, Stratagene

Electrofen-Blue"

/clone\_lib="NIH XGC tropTad5"

/note="Vector: pCS108; Site 1: Sall; Site 2: NotI; Tadpole

library constructed by Russell B. Fletcher in R. Harland's

lab using poly A RNA and oligo dt primers (Invitrogen

SuperScript Plasmid System for cDNA Synthesis and

cloning). Sall (5' end) -NotI (3' end) cDNA was inserted

into vector pCS108

(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

"

Query Match 80.8%; Score 20.2; DB 8; Length 719;

Best Local Similarity 88.0%; Pred. No. 4.3e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
 |||||

Db 558 AAAAAACACCACTGTACTGGG 534  
 |||||

## RESULT 26

CX408645/c

LOCUS

DEFINITION

IMAGE:7607944 5', mRNA sequence.

ACCESSION

CX408645



of California, Berkeley  
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
 Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
<http://image.llnl.gov>  
 Naming Conventions: EST name is generated by the concatenation of  
 the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
 indicates a forward sequencing read of the insert. It does not  
 necessarily reflect the orientation of the insert.  
 Plate: XZG 0429 row: j column: 18  
 High quality sequence stop: 755.  
 Location/Qualifiers

## FEATURES

source

```

1..761
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7556468"
/tissue_type="whole embryo"
/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electrofen-Blue"
/clone_lib="NIH_XGC_tropGas7"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI;
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dT primers
(Invitrogen SuperScript Plasmid System for cDNA Synthesis
and Cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
."

```

## ORIGIN

```

Query Match      80.8%; Score 20.2; DB 8; Length 761;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAACACCACTGAGCTGGG 25
|||||
Db 234 AAAAAACCACTGACTGGG 210
|||||

```

## RESULT 29

```

CF219540
LOCUS CF219540 767 bp mRNA linear EST 04-AUG-2003
DEFINITION AGENCOURT15084592 NICHD_XGC_Emb5 Xenopus tropicalis cDNA clone
IMAGE:6989687 5', mRNA sequence.
CF219540
VERSION CF419540.1 GI:33420248

```

## KEYWORDS

EST.

## SOURCE

```

Xenopus tropicalis (western clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

## REFERENCE

```

AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14664 row: b column: 22
High quality sequence start: 6
High quality sequence stop: 718.
Location/Qualifiers

```

## FEATURES

source

```

1..767
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6989687"
/tissue_type="gastrula"
/dev_stage="embryo, stages 10-13"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_Emb5"
/note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
cloned unidirectionally. Primer: Oligo dT. Average insert
size 2.0 kb. Constructed by Invitrogen. Note: This is a
Xenopus Gene Collection (XGC) library."

```

## ORIGIN

```

Query Match      80.8%; Score 20.2; DB 6; Length 767;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAACACCACTGAGCTGGG 25
|||||
Db 733 AAAAAACCACTGACTGGG 757
|||||

```

## RESULT 30

```

DN099260/c
LOCUS DN099260 768 bp mRNA linear EST 02-AUG-2005
DEFINITION JGI_CABE8370.fwd NIH_XGC_tropoval Xenopus tropicalis cDNA clone
IMAGE:7826648 5', mRNA sequence.
DN099260
VERSION DN099260.1 GI:59769889

```

## KEYWORDS

EST.

## SOURCE

```

Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

## REFERENCE

```

AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other ESTs: JGI_CABE8370.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

```

```

Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Bruce Blumberg Laboratory, University of
California, Irvine
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Plate: CABE 0085 row: d column: 6
High quality sequence stop: 723.
Location/Qualifiers

```

## FEATURES

source

```

1..768
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clone="IMAGE:7826648"
/sex="Female"
/tissue_type="Ovary"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_lib="NIH_XGC_tropOval"

```

/notes="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming (5'-ACTAGTGGCGCCCTAGCCCTGAGTCTTTTCTTTTCTTTT-3') and Stratascript reverse transcriptase. After ligation of EcoRI adapters (5'-AATTCGGCAGCAGG-3') followed by kinasing adapters and by XhoI digestion, the cDNA was size selected by chromatography on Sepharose CL-2B columns and fractions containing cDNAs larger than 1000 bp were ligated into EcoRI/XhoI-digested pCS107. Reference for library construction: Current Genomics 4, 635-644. Library constructed by Michelle Tabb and Bruce Blumberg (Dept of Developmental and Cell Biology, University of California, Irvine)."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 768;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACACACTGAGCTGGG 25  
||||| ||||||| ||||||| |||||||  
Db 673 AAAAAACACACACTGACTGGG 649

## RESULT 31

CX883540/c  
LOCUS  
DEFINITION CX883540 770 bp mRNA linear EST 04-FEB-2005  
IMAGE:7673849 5', mRNA sequence.

ACCESSION CX883540  
VERSION  
SOURCE

KEYWORDS  
SOURCE  
ORGANISM

Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

## REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,  
Brokstein, P. and Lindquist, E.A.  
DOE Joint Genome Institute Xenopus tropicalis EST project  
Unpublished (2004)  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,  
University of California, Berkeley:  
http://tropicalis.berkeley.edu/home)  
cDNA Library Preparation: DOE Joint Genome Institute:  
http://www.jgi.doe.gov  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: CAAL 0233 row: e column: 15  
High quality sequence stop: 761.  
Location/Qualifiers

## FEATURES

source

1..770  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7673849"  
/tissue\_type="Brain"  
/dev\_stage="Adult"  
/lab\_host="Electromax DH10B"  
/clone\_lib="NIH\_XGC\_tropBrn4"

/note="Vector: pCMVSPORT6; Site 1: SalI; Site 2: NotI;  
This library was made from dt primed cDNA and cloned into  
Invitrogen pCMVSPORT6 vector. The work was done at DOE  
Joint Genome Institute. Poly A RNA were primed with 5',  
GACTAGTCTAGATCGCAG CGCGCGCCCTTTTCTTTTCTTTT 3', cDNA  
were ligated to SalI adapter (5'-TCGACCCAGCGTCCG and  
5'-CGACCGCTGGG), digested with NotI, size fractionated in  
1.1% agarose gel electrophoresis and ligated into NotI and  
SalI digested pCMVSPORT6 vector."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 770;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACACACTGAGCTGGG 25  
||||| ||||||| ||||||| |||||||  
Db 277 AAAAAACACACACTGACTGGG 253

## RESULT 32

CX416784/c  
LOCUS  
DEFINITION CX416784 773 bp mRNA linear EST 06-JAN-2005  
IMAGE:7578523 5', mRNA sequence.

ACCESSION CX416784.1 GI:57197487

VERSION

KEYWORDS

SOURCE

ORGANISM

Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

## REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,  
Brokstein, P. and Lindquist, E.A.  
DOE Joint Genome Institute Xenopus tropicalis EST project  
Unpublished (2004)  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Richard M. Harland Laboratory, University of  
California, Berkeley: http://tropicalis.berkeley.edu/home  
cDNA Library Preparation: Richard M. Harland Laboratory, University  
of California, Berkeley  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: XZG 0673 row: a column: 17  
High quality sequence stop: 770.  
Location/Qualifiers

## FEATURES

source

1..773  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7578523"  
/tissue\_type="whole embryo"  
/dev\_stage="Gastrula (st. 10.5-12.5)"  
/lab\_host="E. coli XLI-Blue derivative, Stratagene  
ElectroTen-Blue"  
/clone\_lib="NIH\_XGC\_tropGas7"  
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI;  
Gastrula library constructed by Russell B. Fletcher in R.  
Harland's lab using poly A RNA and oligo dt primers  
(Invitrogen SuperScript Plasmid System for cDNA Synthesis  
and Cloning). SalI (5' end) -NotI (3' end) cDNA was

inserted into vector pCS108  
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)  
."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 773;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAACAACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
DB 172 AAAAACAACCACTGACTGGG 148

## RESULT 33

DR834532  
LOCUS JGI\_CABC5176.fwd NIH XGC tropFat1 Xenopus tropicalis cDNA clone  
DEFINITION IMAGE:7801607 5', mRNA sequence.  
ACCESSION DR834532  
VERSION DR834532.1 GI:71453472  
KEYWORDS EST.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 781)  
Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C.,  
Brokstein,P. and Lindquist,E.A.  
DOE Joint Genome Institute Xenopus tropicalis EST project  
Unpublished (2004)  
Other\_ESTs: JGI\_CABC5176.rev  
Contact: Lindquist,E.A., Richardson,P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: [cdna@jgi-psf.org](mailto:cdna@jgi-psf.org)  
Tissue Procurement: Robert M. Grainger  
cDNA Library Preparation: Bruce Blumberg Laboratory, University of  
California, Irvine  
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
<http://image.llnl.gov>  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: CABC 0053 row: P column: 21  
High quality sequence stop: 701.  
Location/Qualifiers  
1..781  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/strain="N6 (Nigerian 6th generation inbred)"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7801607"  
/tissue\_type="Fat body"  
/dev\_stage="Adult"  
/lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"  
/clone\_lib="NIH XGC tropFat1"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The  
library was prepared from 5 ug of poly A+ RNA by oligo-dT  
priming  
(5'-ACTAGTCGGCGCTAGCGCTCGAGTTTTTTTTTTTTTTT-3') and  
Stratascript reverse transcriptase. After ligation of  
EcoRI adaptors (5'-AATTCGACGAGG-3') followed by kinasing  
adapters and by XhoI digestion, the cDNA was size selected  
by chromatography on Sepharose CL-2B columns and fractions  
containing cDNAs larger than 1000 bp were ligated into  
EcoRI/XhoI-digested pCS107. Reference for library  
construction: Current Genomics 4, 635-644. Library

## FEATURES

source

## ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 781;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAACAACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
DB 717 AAAAACAACCACTGACTGGG 741

## RESULT 34

EX729979/c  
LOCUS BX729979 XGC-tadpole Xenopus tropicalis cDNA clone TTPA025dl5 5',  
DEFINITION mRNA sequence.  
ACCESSION BX729979  
VERSION BX729979.1 GI:38402720  
KEYWORDS EST.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 788)  
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
Unpublished (2003)  
Contact: Croning MDR  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: [trp@sanger.ac.uk](mailto:trp@sanger.ac.uk)  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TTPA025dl5.plkSP6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Nigel Garrett.  
cDNA was oligo dT primed from 5ug of poly A+ RNA from tadpole  
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli DH10B.  
Location/Qualifiers  
1..788  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TTPA025dl5"  
/dev\_stage="tadpole (stage 35-40)"  
/lab\_host="E. coli DH10B"  
/clone\_lib="XGC-tadpole"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from tadpole  
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107  
with EcoRI at the 5' end and NotI at the 3' end"

## FEATURES

source

## ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 788;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAACAACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
DB 515 AAAAACAACCACTGACTGGG 491

## RESULT 35

DN100573  
LOCUS JGI\_CABE9070.fwd NIH XGC tropOval Xenopus tropicalis cDNA clone  
DEFINITION

constructed by Michelle Tabb and Bruce Blumberg (Dept of  
Developmental and Cell Biology, University of California,  
Irvine)."



```

IMAGE:7827590 5', mRNA sequence.
DN100573
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 807)
Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute
Unpublished (2004)
Other ESTs: JGI CABE9070.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Bruce Blumberg Laboratory, University of
California, Irvine
DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: CABE 0093 row: k column: 12
High quality sequence stop: 784.
Location/Qualifiers
1. 807
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clones="IMAGE:7827590"
/sex="Female"
/tissue_type="Ovary"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_lib="NIH_XGC tropOval"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The
library was prepared from 5' ug of poly A+ RNA by oligo-dT
priming
(5'-ACTAGTCGCCGCCCTAGGCTCGAGTTTCTTTTCTTTT-3') and
Stratascript reverse transcriptase. After ligation of
EcoRI adaptors (5'-AATTCGGCAGGAG-3') followed by kinasing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on Sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested pCS107. Reference for library
construction: Current Genomics 4, 635-644. Library
constructed by Michelle Tabb and Bruce Blumberg (Dept of
Developmental and Cell Biology, University of California,
Irvine)."
```

---

```

RESULT 36
CX852874/c
LOCUS
DEFINITION
JGI CAAL9064.fwd NIH_XGC tropBrn4 Xenopus tropicalis cDNA clone
IMAGE:7666788 5', mRNA sequence.
CX852874
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 818)
Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute
Unpublished (2004)
Other ESTs: JGI CAAL9064.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,
University of California, Berkeley:
http://tropicalis.berkeley.edu/home)
cDNA Library Preparation: DOE Joint Genome Institute:
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Plate: CAAL 0093 row: o column: 10
High quality sequence stop: 739.
Location/Qualifiers
1. 818
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clones="IMAGE:7666788"
/tissue_type="Brain"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B"
/clone_lib="NIH_XGC tropBrn4"
/notes="Vector: pCMVSPORT6; Site 1: SalI; Site 2: NotI;
This library was made from dT primed cDNA and cloned into
Invitrogen pCMVSPORT6 vector. The work was done at DOE
Joint Genome Institute. Poly A RNA were primed with 5'
GACTAGTTCTAGATCCGAG CGGCCGCCCTTTTCTTTT 3'. CDNA
were ligated to SalI adapter (5' TCGACCCACGCGCCG and
5'CGGACGGGTGGG), digested with NotI, size fractionated in
1.1% agarose gel electrophoresis and ligated into NotI and
SalI digested pCMVSPORT6 vector."
```

---

```

Query Match 80.8%; Score 20.2; DB 8; Length 818;
Best Local Similarity 88.0%; Pred. NO. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGACTGCTGGG 25
|||||
Db 729 AAAAAACCACTGACTGCTGGG 705
|||||

RESULT 37
CF218992
LOCUS
DEFINITION
AGENCOURT_15085552 NICHD_XGC_Emb5 Xenopus tropicalis cDNA clone
IMAGE:7666788 5', mRNA sequence.
CF218992
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 820)
Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute
Unpublished (2004)
Other ESTs: JGI CABE9070.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Bruce Blumberg Laboratory, University of
California, Irvine
DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: CABE 0093 row: k column: 12
High quality sequence stop: 784.
Location/Qualifiers
1. 820
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clones="IMAGE:7827590"
/sex="Female"
/tissue_type="Ovary"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_lib="NIH_XGC tropOval"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The
library was prepared from 5' ug of poly A+ RNA by oligo-dT
priming
(5'-ACTAGTCGCCGCCCTAGGCTCGAGTTTCTTTTCTTTT-3') and
Stratascript reverse transcriptase. After ligation of
EcoRI adaptors (5'-AATTCGGCAGGAG-3') followed by kinasing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on Sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested pCS107. Reference for library
construction: Current Genomics 4, 635-644. Library
constructed by Michelle Tabb and Bruce Blumberg (Dept of
Developmental and Cell Biology, University of California,
Irvine)."
```

---

```

Query Match 80.8%; Score 20.2; DB 8; Length 807;
Best Local Similarity 88.0%; Pred. NO. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGACTGCTGGG 25
|||||
Db 86 AAAAAACCACTGACTGCTGGG 62
|||||
```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	IMAGE:6989266 5', mRNA sequence.	
	CF218992	
	CF218992.1 GI:33419700	
	EST.	
REFERENCE AUTHORS TITLE JOURNAL COMMENT	Xenopus tropicalis (western clawed frog)	
	Xenopus tropicalis	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.	
	1 (bases 1 to 820)	
FEATURES source	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	
	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: <a href="mailto:cgabbs-r@mail.nih.gov">cgabbs-r@mail.nih.gov</a> Tissue Procurement: Robert M. Grainger CDNA Library Prepared by: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLNL4663 row: a column: 09 High quality sequence start: 6 High quality sequence stop: 735. Location/Qualifiers 1..820 /organism="Xenopus tropicalis" /mol_type="mRNA" /db_xref="taxon:8364" /clone="IMAGE:6989266" /tissue_type="gastrula" /dev_stage="embryo, stages 10-13" /lab_host="DH10B (phage-resistant)" /clone_lib="NICHED_XGC_Emb5" /notes="Vector: pCMV-SF0K16.1; Site_1: NotI; Site_2: EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.0 kb. Constructed by Invitrogen. Note: This is a Xenopus Gene Collection (XGC) library."	
ORIGIN	Query Match 80.8%; Score 20.2; DB 6; Length 820; Best Local Similarity 88.0%; Pred. No. 4.3e+02; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	QY 1 AAAAAACCAACCACTGAGCTGGG 25	
	Db 750 AAAAAACCAACCACTGACTGGG 774	
ORIGIN	RESULT 39	
	CX403435/c	
	LOCUS	
	DEFINITION	
ORIGIN	ACCESSION	
	VERSION	
	KEYWORDS	
	SOURCE	
ORIGIN	Xenopus tropicalis (western clawed frog)	
	Xenopus tropicalis	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.	
	1 (bases 1 to 822)	
ORIGIN	Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.	
	DOE Joint Genome Institute	
	Unpublished (2004)	
	Other ESTs: JGI XZT50439.rev	
ORIGIN	Contact: Lindquist, E.A., Richardson, P.	
	DOE Joint Genome Institute	
	2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
	Tel: 925 296 5600	
ORIGIN	Fax: 925 296 5710	
	Email: <a href="mailto:cdna@jgi-psf.org">cdna@jgi-psf.org</a>	
	Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <a href="http://tropicalis.berkeley.edu/home">http://tropicalis.berkeley.edu/home</a>	
	CDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley	
ORIGIN	DNA Sequencing: DOE Joint Genome Institute: <a href="http://www.jgi.doe.gov">http://www.jgi.doe.gov</a>	
	Clone Distribution: I.M.A.G.E. Consortium/LLNL: <a href="http://image.llnl.gov">http://image.llnl.gov</a>	
	Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.	
	Plate: XZT 0525 row: n column: 9	

High quality sequence stop: 748.

# FEATURES

source  
1..822  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7625675"  
/tissue\_type="whole embryo"  
/dev\_stage="radpole (st. 36-41)"  
/lab\_host="E. coli XL1-Blue derivative, Stratagene Electropen-Blue"  
/clone\_lib="NIH XGC tropTad5"  
/note="vector: pCS107; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS107  
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 822;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||||| ||||||| |||||||  
Db 624 AAAAAACCACTGAGCTGGG 600

# RESULT 40

CR444938 830 bp mRNA linear EST 19-JUN-2004  
LOCUS CR444938 XGC-tailbud Xenopus tropicalis cDNA clone TTBA065f19 3',  
DEFINITION mRNA sequence.  
ACCESSION CR444938.1 GI:48970525  
VERSION Est.  
KEYWORDS Xenopus tropicalis (western clawed frog)  
SOURCE Xenopus tropicalis  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.  
REFERENCE 1 (bases 1 to 830)  
Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.  
Sanger Xenopus tropicalis EST project 2001 (2004)  
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)  
JOURNAL Unpublished (2004)  
COMMENT Contact: Croning MDR  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TTBA065f19.q1kT7  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.  
Seq primer: T7.

# FEATURES

source  
1..830  
Location/Qualifiers  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TTBA065f19"  
/dev\_stage="tailbud (stage 28-30)"  
/lab\_host="Escherichia coli DH10B."  
/clone\_lib="XGC-tailbud"  
/note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 830;

Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

# Qy

1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||||| ||||||| |||||||

# Db

767 AAAAAACCACTGAGCTGGG 791  
||||| ||||||| ||||||| |||||||

# RESULT 41

DR863945 836 bp mRNA linear EST 29-JUL-2005  
LOCUS JGI\_CABG7661.fwd NIH XGC\_tropStol Xenopus tropicalis cDNA clone  
DEFINITION IMAGE:7838710 5', mRNA sequence.  
ACCESSION DR863945  
VERSION DR863945.1 GI:71520535  
KEYWORDS EST.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.  
REFERENCE 1 (bases 1 to 836)  
AUTHORS Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.  
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project  
JOURNAL Unpublished (2004)  
COMMENT Other ESTs: JGI CABG7661.rev  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-bef.org  
Tissue Procurement: Robert M. Grainger  
CDNA Library Preparation: Bruce Blumberg Laboratory, University of California, Irvine  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
http://image.llnl.gov  
Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.  
Plate: CABG 0077 row: j column: 20  
High quality sequence stop: 799.

# FEATURES

source  
1..836  
Location/Qualifiers  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/strain="N6 (Nigerian 6th generation inbred)"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7838710"  
/tissue\_type="Stomach"  
/dev\_stage="Adult"  
/lab\_host="NIH XGC tropStol"  
/clone\_lib="ElectroMAX DH10B T1 Phage Resistant cells"  
/note="vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dt priming  
(5'-ACTAGTCGCGCGCTAGCCTCGAGTTTTTTTTTTTTTTTTTTT-3') and Stratascript reverse transcriptase. After ligation of EcoRI adapters (5'-AATCGCAGCAGG-3') followed by kinasing adapters and by XhoI digestion, the cDNA was size selected by chromatography on Sepharose CL-2B columns and fractions containing cDNAs larger than 1000 bp were ligated into EcoRI/XhoI-digested pCS107. Reference for library construction: Current Genomics 4, 635-644. Library constructed by Michelle Tabb and Bruce Blumberg (Dept of Developmental and Cell Biology, University of California, Irvine)."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 836;

Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
Db 777 AAAAAACCACTGACTGGG 801

RESULT 42  
DR863944/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus tropicalis (western clawed frog)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 838)  
Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C.,  
Brokstein,P. and Lindquist,E.A.  
DOE Joint Genome Institute  
Unpublished (2004)  
Other ESTs: JGI\_CABG7661.fwd  
Contact: Lindquist,E.A., Richardson,P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 236 5600  
Fax: 925 236 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Robert M. Grainger  
DNA Library Preparation: Bruce Blumberg Laboratory, University of  
California, Irvine  
Cloning: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LNL:  
<http://image.lnl.gov>  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.rev'  
indicates a reverse sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: CABG 0077 row: j column: 20  
High quality sequence stop: 820.  
Location/Qualifiers  
1. .838  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/strain="N6 (Nigerian 6th generation inbred)"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7838710"  
/tissue\_type="Stomach"  
/dev\_stage="Adult"  
/lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"  
/clone\_lib="NIH\_XGC\_tropStol"  
/notes="Vector: PCS107; Site 1: EcoRI; Site 2: XhoI; The  
library was prepared from 5 ug of poly A+ RNA by oligo-dT  
priming  
(5'-ACTAGTCGGCGCTAGCCTCGAGTTTCTTTTCTTTTCTTTT-3') and  
Stratascript reverse transcriptase. After ligation of  
EcoRI adapters (5'-AATTCGACGAGG-3') followed by kinaasing  
adapters and by XhoI digestion, the cDNA was size selected  
by chromatography on Sepharose CL-2B columns and fractions  
containing cDNAs larger than 1000 bp were ligated into  
EcoRI/XhoI-digested PCS107. Reference for library  
construction: Current Genomics 4, 635-644. Library  
constructed by Michelle Tabb and Bruce Blumberg (Dept of  
Developmental and Cell Biology, University of California,  
Irvine)."

Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
Db 777 AAAAAACCACTGACTGGG 801

RESULT 42  
DR863944/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus tropicalis (western clawed frog)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 838)  
Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C.,  
Brokstein,P. and Lindquist,E.A.  
DOE Joint Genome Institute  
Unpublished (2004)  
Other ESTs: JGI\_CABG7661.fwd  
Contact: Lindquist,E.A., Richardson,P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 236 5600  
Fax: 925 236 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Robert M. Grainger  
DNA Library Preparation: Bruce Blumberg Laboratory, University of  
California, Irvine  
Cloning: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LNL:  
<http://image.lnl.gov>  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.rev'  
indicates a reverse sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Plate: CABG 0077 row: j column: 20  
High quality sequence stop: 820.  
Location/Qualifiers  
1. .838  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/strain="N6 (Nigerian 6th generation inbred)"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7838710"  
/tissue\_type="Stomach"  
/dev\_stage="Adult"  
/lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"  
/clone\_lib="NIH\_XGC\_tropStol"  
/notes="Vector: PCS107; Site 1: EcoRI; Site 2: XhoI; The  
library was prepared from 5 ug of poly A+ RNA by oligo-dT  
priming  
(5'-ACTAGTCGGCGCTAGCCTCGAGTTTCTTTTCTTTTCTTTT-3') and  
Stratascript reverse transcriptase. After ligation of  
EcoRI adapters (5'-AATTCGACGAGG-3') followed by kinaasing  
adapters and by XhoI digestion, the cDNA was size selected  
by chromatography on Sepharose CL-2B columns and fractions  
containing cDNAs larger than 1000 bp were ligated into  
EcoRI/XhoI-digested PCS107. Reference for library  
construction: Current Genomics 4, 635-644. Library  
constructed by Michelle Tabb and Bruce Blumberg (Dept of  
Developmental and Cell Biology, University of California,  
Irvine)."

Query Match 80.8%; Score 20.2; DB 8; Length 838;

Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
Db 236 AAAAAACCACTGACTGGG 212

RESULT 43  
BX782362  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 844)  
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
Unpublished (2003)  
Contact: Croning MDR  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TEGG069p15.q1kt7  
Sequencing primer: T7  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
cDNA was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the  
5' end and NotI at the 3' end.  
Vector: PCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli XLI-blue.  
Location/Qualifiers  
1. .844  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TEGG069p15"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XLI-blue"  
/clone\_lib="XGC-egg"  
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into PCS107 with  
EcoRI at the 5' end and NotI at the 3' end"

FEATURES  
source  
1. .844  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TEGG069p15"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XLI-blue"  
/clone\_lib="XGC-egg"  
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into PCS107 with  
EcoRI at the 5' end and NotI at the 3' end"

Query Match 80.8%; Score 20.2; DB 5; Length 844;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
Db 781 AAAAAACCACTGACTGGG 805

RESULT 44  
CR423927  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 844)  
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
Unpublished (2003)  
Contact: Croning MDR  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TEGG069p15.q1kt7  
Sequencing primer: T7  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
cDNA was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the  
5' end and NotI at the 3' end.  
Vector: PCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli XLI-blue.  
Location/Qualifiers  
1. .844  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TEGG069p15"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XLI-blue"  
/clone\_lib="XGC-egg"  
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dT primed from 5ug of poly A+ RNA from egg.  
EcoRI-NotI cut cDNA was then ligated into PCS107 with  
EcoRI at the 5' end and NotI at the 3' end"

Query Match 80.8%; Score 20.2; DB 5; Length 844;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| ||||| |||||  
Db 781 AAAAAACCACTGACTGGG 805

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 848)

CR441514 XGC-tailbud Xenopus tropicalis EST project 2001 (2004)

Unpublished (2004)

Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TTBA013109.q1kT7

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Nigel Garrett.

Seq primer: T7.

Location/Qualifiers

1..848

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TTBA013109"

/dev\_stage="tailbud (stage 28-30)"

/lab\_host="Escherichia coli DH10B."

/clone\_lib="XGC-tailbud"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from 5' of poly A+ RNA from tailbud.

EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end."

#### ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 7; Length 848;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

||||| ||||| ||||| ||||| |||||

DB 788 AAAAAACCACTGACTGGG 812

RESULT 45

CR441514

LOCUS

CR441514 XGC-tailbud Xenopus tropicalis cDNA clone TTBA076108 3',

mRNA sequence.

CR441514

CR441514.1 GI:48967101

EST.

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 851)

Croning M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (2004)

Unpublished (2004)

Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TTBA076108.q1kT7

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Nigel Garrett.

Seq primer: T7.

Location/Qualifiers

1..851

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TTBA076108"

/dev\_stage="tailbud (stage 28-30)"

/lab\_host="Escherichia coli DH10B."

/clone\_lib="XGC-tailbud"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from 5' of poly A+ RNA from tailbud.

EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end."

#### ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 7; Length 851;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

||||| ||||| ||||| ||||| |||||

DB 788 AAAAAACCACTGACTGGG 812

RESULT 46

BX771898

LOCUS

BX771898 XGC-egg Xenopus tropicalis cDNA clone Tegg044013 3', mRNA

sequence.

BX771898

EST.

Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 862)

Croning M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: Tegg044013.q1kT7

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from 5' of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the

5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-Blue.

Location/Qualifiers

1..862

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="Tegg044013"

/dev\_stage="egg"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-egg"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from 5' of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end"

#### ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 5; Length 862;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

||||| ||||| ||||| ||||| |||||

DB 782 AAAAAACCACTGACTGGG 806

RESULT 47

```
CR568362          862 bp  mRNA  linear  EST 19-JUL-2004
LOCUS             3', mRNA sequence.
DEFINITION        CR568362 XGC-tailbud-head Xenopus tropicalis cDNA clone THda017b03
ACCESSION         CR568362
VERSION           CR568362.1 GI:50398439
KEYWORDS          EST.
SOURCE            Xenopus tropicalis (western clawed frog)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
                  Xenopodinae; Xenopus; Silurana.
REFERENCE         1 (bases 1 to 862)
AUTHORS           Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
TITLE             Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL           Unpublished (2004)
COMMENT           Contact: Croning MDR
                  Sanger Institute
                  Hinxton, Cambridgeshire, CB10 1SA, UK
                  Email: trop@sanger.ac.uk
                  Sanger Xenopus tropicalis EST project 2001
                  TROPICALIS SEQUENCE ID: THda017b03.q1kat7
                  This sequence is from a Xenopus Gene Collection (XGC) library
                  constructed by Nigel Garrett.
                  Seq primer: T7
                  Location/Qualifiers
                    1..862
                    /organism="Xenopus tropicalis"
                    /mol_type="mRNA"
                    /db_xref="taxon:8364"
                    /clone="THda017b03"
                    /dev_stage="tailbud head (stage 28-30)"
                    /lab_host="Escherichia coli DH10B."
                    /clone_lib="XGC-tailbud-head"
                    /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from tailbud
head. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end."
ORIGIN
Query Match      80.8%; Score 20.2; DB 7; Length 862;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACTGAGCTGGG 25
    ||||| ||||| ||||| |||||
Db 793 AAAAAAACCACTGACTGGG 817

RESULT 48
BX755834          864 bp  mRNA  linear  EST 10-DEC-2003
LOCUS             mRNA sequence.
DEFINITION        BX755834 XGC-gastrula Xenopus tropicalis cDNA clone TGas116g11 3',
ACCESSION         BX755834
VERSION           BX755834.1 GI:39663042
KEYWORDS          EST.
SOURCE            Xenopus tropicalis (western clawed frog)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
                  Xenopodinae; Xenopus; Silurana.
REFERENCE         1 (bases 1 to 864)
AUTHORS           Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE             Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL           Unpublished (2003)
COMMENT           Contact: Croning MDR
                  Sanger Institute
                  Hinxton, Cambridgeshire, CB10 1SA, UK
                  Email: trop@sanger.ac.uk
                  Sanger Xenopus tropicalis EST project 2001
                  TROPICALIS SEQUENCE ID: TGas116g11.q1kat7
                  Sequencing primer: T7
```

```
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XLI-blue.

FEATURES
source
Location/Qualifiers
  1..864
  /organism="Xenopus tropicalis"
  /mol_type="mRNA"
  /db_xref="taxon:8364"
  /clone="TGas116g11"
  /dev_stage="gastrula (stages 10.5-12 mixed)"
  /lab_host="Escherichia coli XLI-blue"
  /clone_lib="XGC-gastrula"
  /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
ORIGIN
Query Match      80.8%; Score 20.2; DB 5; Length 864;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAACCACTGAGCTGGG 25
    ||||| ||||| ||||| |||||
Db 801 AAAAAACCACTGACTGGG 825

RESULT 49
BX750203          869 bp  mRNA  linear  EST 09-DEC-2003
LOCUS             mRNA sequence.
DEFINITION        BX750203 XGC-gastrula Xenopus tropicalis cDNA clone TGas086a19 3',
ACCESSION         BX750203
VERSION           BX750203.1 GI:39640271
KEYWORDS          EST.
SOURCE            Xenopus tropicalis (western clawed frog)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
                  Xenopodinae; Xenopus; Silurana.
REFERENCE         1 (bases 1 to 869)
AUTHORS           Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE             Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL           Unpublished (2003)
COMMENT           Contact: Croning MDR
                  Sanger Institute
                  Hinxton, Cambridgeshire, CB10 1SA, UK
                  Email: trop@sanger.ac.uk
                  Sanger Xenopus tropicalis EST project 2001
                  TROPICALIS SEQUENCE ID: TGas086a19.q1kat7
                  Sequencing primer: T7
                  This sequence is from a Xenopus Gene Collection (XGC) library
                  constructed by Aaron M. Zorn.
                  cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
                  gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
                  EcoRI at the 5' end and NotI at the 3' end.
                  Vector: pCS107; Site 1: EcoRI; Site 2: NotI
                  Host: Escherichia coli XLI-blue.
                  Location/Qualifiers
                    1..869
                    /organism="Xenopus tropicalis"
                    /mol_type="mRNA"
                    /db_xref="taxon:8364"
                    /clone="TGas086a19"
                    /dev_stage="gastrula (stages 10.5-12 mixed)"
                    /lab_host="Escherichia coli XLI-blue"
                    /clone_lib="XGC-gastrula"
                    /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
FEATURES
source
Location/Qualifiers
  1..869
  /organism="Xenopus tropicalis"
  /mol_type="mRNA"
  /db_xref="taxon:8364"
  /clone="TGas086a19"
  /dev_stage="gastrula (stages 10.5-12 mixed)"
  /lab_host="Escherichia coli XLI-blue"
  /clone_lib="XGC-gastrula"
  /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
```

Mon Feb 6 12:23:17 2006

was oligo dT primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 869;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| |||||  
Db 788 AAAAAACCACTGACTGGG 812

RESULT 50

CR568361/c  
LOCUS CR568361 XGC-tailbud-head Xenopus tropicalis cDNA clone THdA017b03  
DEFINITION 872 bp mRNA linear EST 19-JUL-2004  
S' mRNA sequence.

ACCESSION CR568361  
VERSION CR568361.1 GI:50398438

KEYWORDS EST.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 872)  
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.  
TITLE Sahger Xenopus tropicalis EST project 2001 (2004)  
JOURNAL Unpublished (2004)  
COMMENT Contact: Croning MDR

Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: THdA017b03.pikasp6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Nigel Garrett.  
Seq primer: SP6.

FEATURES

source  
1..872 Location/Qualifiers  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="THdA017b03"  
/dev\_stage="tailbud head (stage 28-30)"  
/lab\_host="Escherichia coli DH10B."  
/clone\_lib="XGC-tailbud-head"  
/note="vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dT primed from Sug of poly A+ RNA from tailbud  
head. EcoRI-NotI cut cDNA was then ligated into pCS107  
with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 872;  
Best Local Similarity 88.0%; Pred. No. 4.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACCACTGAGCTGGG 25  
||||| ||||| ||||| |||||  
Db 713 AAAAAACCACTGACTGGG 689

Search completed: February 3, 2006, 22:01:45  
Job time : 2961.67 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-3  
Perfect score: 25  
Sequence: 1 aaaaacaccaccactgagctggg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues  
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*

7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*

8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*

9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.6	74.4	6569	3	US-09-949-016-14892
2	18.6	74.4	6758	3	US-09-949-016-16871
3	18.6	74.4	15593	3	US-09-949-016-17177
4	18.6	74.4	26843	3	US-09-949-016-17208
5	18.6	74.4	106924	3	US-09-949-016-13834
6	18.2	72.8	407	3	US-09-513-999C-26191
7	18.2	72.8	601	3	US-09-949-016-69126
8	18.2	72.8	601	3	US-09-949-016-70677
9	18.2	72.8	601	3	US-09-949-016-70678
10	18.2	72.8	601	3	US-09-949-016-159778
11	18.2	72.8	601	3	US-09-949-016-195472
12	18.2	72.8	601	3	US-09-949-016-195473
13	18.2	72.8	54246	3	US-09-949-016-16206
14	18.2	72.8	58844	3	US-09-949-016-13769
15	18.2	72.8	121114	3	US-09-949-016-17232
16	18.2	72.8	154605	3	US-09-949-016-11894
17	18.2	72.8	455726	3	US-09-949-016-14157
18	18.2	72.8	481115	3	US-09-949-016-11940
19	17.8	71.2	1448	3	US-08-858-207A-71
20	17.8	71.2	1509	3	US-08-489-039A-5144
21	17.8	71.2	1542	2	US-08-865-311-1
22	17.8	71.2	1542	3	US-09-315-720-1
23	17.8	71.2	1545	3	US-09-583-110-1228
24	17.8	71.2	1566	3	US-09-107-433-1292

Sequence 7, Appli	US-08-961-527-7	17.2	19702	3	Sequence 16822, A
Sequence 16822, A	US-09-949-016-16822	71.2	73818	3	Sequence 24377, A
Sequence 24377, A	US-09-949-016-24377	70.4	601	3	Sequence 134073, A
Sequence 134073, A	US-09-949-016-134073	70.4	601	3	Sequence 199548, A
Sequence 199548, A	US-09-949-016-199548	70.4	601	3	Sequence 15944, A
Sequence 15944, A	US-09-949-016-15944	70.4	93532	3	Sequence 12173, A
Sequence 12173, A	US-09-949-016-12173	70.4	221958	3	Sequence 15498, A
Sequence 15498, A	US-09-949-016-15498	70.4	221966	3	Sequence 17371, A
Sequence 17371, A	US-09-949-016-17371	70.4	373182	3	Sequence 12062, A
Sequence 12062, A	US-09-949-016-12062	70.4	373694	3	Sequence 3918, Ap
Sequence 3918, Ap	US-09-248-796A-3918	69.6	705	3	Sequence 1118, Ap
Sequence 1118, Ap	US-09-248-796A-1118	69.6	882	3	Sequence 15984, A
Sequence 15984, A	US-09-949-016-15984	69.6	22581	3	Sequence 15955, A
Sequence 15955, A	US-09-949-016-15954	69.6	54576	3	Sequence 15955, A
Sequence 15955, A	US-09-949-016-15955	69.6	54576	3	Sequence 13734, A
Sequence 13734, A	US-09-949-016-13734	69.6	131254	3	Sequence 22643, A
Sequence 22643, A	US-09-949-016-22643	68.8	601	3	Sequence 36627, A
Sequence 36627, A	US-09-949-016-36627	68.8	601	3	Sequence 44701, A
Sequence 44701, A	US-09-949-016-44701	68.8	601	3	Sequence 59992, A
Sequence 59992, A	US-09-949-016-59992	68.8	601	3	Sequence 172749, A
Sequence 172749, A	US-09-949-016-172749	68.8	601	3	Sequence 184440, A
Sequence 184440, A	US-09-949-016-184440	68.8	862	3	Sequence 2205, Ap
Sequence 2205, Ap	US-09-533-559-2205	68.8	862	3	Sequence 4, Appli
Sequence 4, Appli	US-08-991-946A-4	68.8	1543	2	Sequence 953, App
Sequence 953, App	US-09-949-016-953	68.8	2103	3	Sequence 1278, Ap
Sequence 1278, Ap	US-09-949-016-1278	68.8	2103	3	Sequence 12695, A
Sequence 12695, A	US-09-949-016-12695	68.8	7659	3	Sequence 13020, A
Sequence 13020, A	US-09-949-016-13020	68.8	7665	3	Sequence 11971, A
Sequence 11971, A	US-09-949-016-11971	68.8	30973	3	Sequence 16628, A
Sequence 16628, A	US-09-949-016-16628	68.8	30974	3	Sequence 17339, A
Sequence 17339, A	US-09-949-016-17338	68.8	60110	3	Sequence 14144, A
Sequence 14144, A	US-09-949-016-17339	68.8	60110	3	Sequence 14145, A
Sequence 14145, A	US-09-949-016-14144	68.8	61083	3	Sequence 12174, A
Sequence 12174, A	US-09-949-016-14145	68.8	61083	3	Sequence 15635, A
Sequence 15635, A	US-09-949-016-12174	68.8	153642	3	Sequence 12707, A
Sequence 12707, A	US-09-949-016-15635	68.8	153643	3	Sequence 17026, A
Sequence 17026, A	US-09-949-016-17026	68.8	190078	3	Sequence 15654, A
Sequence 15654, A	US-09-949-016-15654	68.8	190078	3	Sequence 13498, A
Sequence 13498, A	US-09-949-016-13498	68.8	343352	3	Sequence 1, Appli
Sequence 1, Appli	US-08-545-528D-1	68.8	580073	3	Sequence 21681, A
Sequence 21681, A	US-09-513-999C-21681	68.0	208	3	Sequence 54261, A
Sequence 54261, A	US-09-949-016-54261	68.0	601	3	Sequence 54262, A
Sequence 54262, A	US-09-949-016-54262	68.0	601	3	Sequence 54325, A
Sequence 54325, A	US-09-949-016-54326	68.0	601	3	Sequence 54326, A
Sequence 54326, A	US-09-949-016-54326	68.0	601	3	Sequence 87785, A
Sequence 87785, A	US-09-949-016-87785	68.0	601	3	Sequence 160657, A
Sequence 160657, A	US-09-949-016-160657	68.0	601	3	Sequence 183145, A
Sequence 183145, A	US-09-949-016-160658	68.0	601	3	Sequence 183886, A
Sequence 183886, A	US-09-949-016-183145	68.0	601	3	Sequence 5163, Ap
Sequence 5163, Ap	US-09-533-559-5163	68.0	639	3	Sequence 1359, Ap
Sequence 1359, Ap	US-10-104-047-1359	68.0	1558	3	Sequence 25, Appl
Sequence 25, Appl	US-09-227-595-25	68.0	1747	3	Sequence 27, Appl
Sequence 27, Appl	US-09-227-595-27	68.0	1747	3	Sequence 25, Appl
Sequence 25, Appl	US-08-595-590B-25	68.0	1747	3	Sequence 43, Appl
Sequence 43, Appl	US-08-595-590B-27	68.0	1747	3	Sequence 1, Appli
Sequence 1, Appli	US-07-916-098A-43	68.0	2029	2	Sequence 4, Appli
Sequence 4, Appli	US-08-070-116A-1	68.0	2399	2	Sequence 14377, A
Sequence 14377, A	US-07-916-098A-44	68.0	2399	2	Sequence 41, Appl
Sequence 41, Appl	US-09-949-016-14377	68.0	9931	3	Sequence 41, Appl
Sequence 41, Appl	US-08-444-644-41	68.0	10844	3	Sequence 14378, A
Sequence 14378, A	US-08-232-246A-41	68.0	10844	3	Sequence 16359, A
Sequence 16359, A	US-09-949-016-14378	68.0	11194	3	Sequence 15783, A
Sequence 15783, A	US-09-949-016-16359	68.0	12787	3	Sequence 1688, App
Sequence 1688, App	US-09-949-016-15783	68.0	14189	3	Sequence 16716, A
Sequence 16716, A	US-09-949-016-16889	68.0	18900	3	Sequence 3, Appli
Sequence 3, Appli	US-09-949-002-668	68.0	47677	3	Sequence 3, Appli
Sequence 3, Appli	US-09-949-016-16716	68.0	54986	3	Sequence 3, Appli
Sequence 3, Appli	US-09-884-890-3	68.0	75395	3	Sequence 3, Appli
Sequence 3, Appli	US-10-274-197-3	68.0	75395	3	Sequence 3, Appli
Sequence 3, Appli	US-10-760-407-3	68.0	75395	3	

```
98 17 68.0 78530 3 US-09-949-016-16790 Sequence 16790, A
99 17 68.0 95561 3 US-09-949-016-12768 Sequence 12768, A
100 17 68.0 95561 3 US-09-949-016-13306 Sequence 13306, A
101 17 68.0 95561 3 US-09-949-016-13307 Sequence 13307, A
102 17 68.0 102304 3 US-09-949-016-12589 Sequence 12589, A
103 17 68.0 107820 3 US-09-792-616-1 Sequence 1, Appli
104 17 68.0 110243 3 US-09-949-016-13698 Sequence 13698, A
105 17 68.0 111424 3 US-09-949-016-17014 Sequence 17014, A
106 17 68.0 112874 3 US-09-949-016-13180 Sequence 13180, A
107 17 68.0 151088 3 US-09-949-016-16240 Sequence 16240, A
108 17 68.0 190078 3 US-09-949-016-12707 Sequence 12707, A
109 17 68.0 190078 3 US-09-949-016-17026 Sequence 17026, A
110 17 68.0 236474 3 US-09-949-016-13418 Sequence 13418, A
111 17 68.0 421118 3 US-09-949-016-16297 Sequence 16297, A
112 17 68.0 670689 3 US-09-949-016-12505 Sequence 12505, A
113 17 68.0 670689 3 US-09-949-016-12505 Sequence 12505, A
114 16.8 67.2 178 3 US-09-621-976-17983 Sequence 17983, A
115 16.8 67.2 601 3 US-09-949-016-55149 Sequence 55149, A
116 16.8 67.2 601 3 US-09-949-016-62603 Sequence 62603, A
117 16.8 67.2 601 3 US-09-949-016-62604 Sequence 62604, A
118 16.8 67.2 601 3 US-09-949-016-165370 Sequence 165370, A
119 16.8 67.2 601 3 US-09-949-016-179675 Sequence 179675, A
120 16.8 67.2 601 3 US-09-949-016-179676 Sequence 179676, A
121 16.8 67.2 669 4 US-09-605-703B-2885 Sequence 2885, Ap
122 16.8 67.2 790 3 US-10-170-098A-1 Sequence 1, Appli
123 16.8 67.2 7822 3 US-09-949-016-13339 Sequence 13339, A
124 16.8 67.2 16914 3 US-09-949-016-15972 Sequence 15972, A
125 16.8 67.2 48940 3 US-09-949-016-16402 Sequence 16402, A
126 16.8 67.2 58361 3 US-09-949-016-16755 Sequence 16755, A
127 16.8 67.2 58361 3 US-09-949-016-16756 Sequence 16756, A
128 16.8 67.2 59065 3 US-09-813-817-3 Sequence 3, Appli
129 16.8 67.2 59065 3 US-09-978-197-3 Sequence 3, Appli
130 16.8 67.2 59065 3 US-10-135-696-3 Sequence 3, Appli
131 16.8 67.2 59065 3 US-10-820-230-3 Sequence 3, Appli
132 16.8 67.2 71815 3 US-09-949-016-12501 Sequence 12501, A
133 16.8 67.2 163662 3 US-09-949-016-12545 Sequence 12545, A
134 16.8 67.2 163662 3 US-09-949-016-13546 Sequence 13546, A
135 16.8 67.2 390416 3 US-09-949-016-15923 Sequence 15923, A
136 16.6 66.4 348 3 US-09-023-655-632 Sequence 632, App
137 16.6 66.4 399 3 US-09-621-976-10504 Sequence 10504, A
138 16.6 66.4 451 3 US-09-270-767-8403 Sequence 8403, Ap
139 16.6 66.4 451 3 US-09-270-767-23685 Sequence 23685, A
140 16.6 66.4 601 3 US-09-949-016-34548 Sequence 34548, A
141 16.6 66.4 601 3 US-09-949-016-41603 Sequence 41603, A
142 16.6 66.4 601 3 US-09-949-016-56245 Sequence 56245, A
143 16.6 66.4 601 3 US-09-949-016-56246 Sequence 56246, A
144 16.6 66.4 601 3 US-09-949-016-60495 Sequence 60495, A
145 16.6 66.4 601 3 US-09-949-016-78004 Sequence 78004, A
146 16.6 66.4 601 3 US-09-949-016-87388 Sequence 87388, A
147 16.6 66.4 601 3 US-09-949-016-147198 Sequence 147198,
148 16.6 66.4 1129 3 US-09-799-451-28 Sequence 28, Appl
149 16.6 66.4 1363 3 US-09-023-655-907 Sequence 907, App
150 16.6 66.4 1523 3 US-09-019-095A-21 Sequence 21, Appl
```

## ALIGNMENTS

```
RESULT 1
US-09-949-016-14892
; Sequence 14892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; SOFTWARE: FastSeq for Windows Version 4.0
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-20
; SOFTWARE: FastSeq for Windows Version 4.0
; PRIOR FILING DATE: 2000-10-20
; SEQ ID NO 17177

Query Match 74.4%; Score 18.6; DB 3; Length 6569;
Best Local Similarity 84.0%; Pred. No. le+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCTGGG 25
Db 6399 AAAAAACAACCACTGAGCTGGG 6423

RESULT 2
US-09-949-016-16871
; Sequence 16871, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16871
; LENGTH: 6758
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16871

Query Match 74.4%; Score 18.6; DB 3; Length 6758;
Best Local Similarity 84.0%; Pred. No. le+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCTGGG 25
Db 6597 AAAAAACAACCACTGAGCTGGG 6621

RESULT 3
US-09-949-016-17177/c
; Sequence 17177, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; PRIOR FILING DATE: 2000-10-20
; SEQ ID NO 17177
```

```
; LENGTH: 15593
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17177

Query Match          74.4%; Score 18.6; DB 3; Length 15593;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 4946 AAGAAAGAACTCCCTGAGCTGGG 4922

RESULT 4
US-09-949-016-17208/c
; Sequence 17208, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17208
; LENGTH: 26843
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17208

Query Match          74.4%; Score 18.6; DB 3; Length 26843;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 5142 AAAAAACACCACCTGAGCTGTG 5118

RESULT 5
US-09-949-016-13834/c
; Sequence 13834, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13834
; LENGTH: 106924
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: {1}...(106924)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13834

Query Match          74.4%; Score 18.6; DB 3; Length 106924;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 4267 AAAAAACACCACACAGAGTAGG 4243

RESULT 6
US-09-513-999C-26191/c
; Sequence 26191, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 26191
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 339
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 382
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 390
; OTHER INFORMATION: s=g or c
US-09-513-999C-26191

Query Match          72.8%; Score 18.2; DB 3; Length 407;
Best Local Similarity 87.0%; Pred. No. 99;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTG 23
   ||||| ||||| ||||| ||||| |||||
Db 36 AAAAAACAGACCACTGATCTG 14

RESULT 7
US-09-949-016-69126/c
; Sequence 69126, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69126
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-69126

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACCACTGAGCTGG 24
    ||||| ||| ||| |||||
DB 450 AAAAACACCACTGAGCTGG 428

RESULT 8
US-09-949-016-70677/c
; Sequence 70677, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70677
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-70677

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACCACTGAGCTGG 25
    ||||| ||| ||| |||||
DB 309 AAAAACACCACTGAGCTGG 285

RESULT 9
US-09-949-016-70678/c
; Sequence 70678, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70678
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-70678

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACCACTGAGCTGG 23
    ||||| ||| ||| |||||
DB 379 AAAAACACCACTGAGCTGG 357

RESULT 11
US-09-949-016-195472
; Sequence 195472, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195472
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195472

Query Match          72.8%; Score 18.2; DB 3; Length 601;
```

```
Best Local Similarity 87.0%; Pred. No. 1.1e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 2 AAAAACACACCCACTGAGCTGG 24
|||||
Db 245 AAAAACACACCCACAGGTTGG 267

RESULT 12
US-09-949-016-195473
; Sequence 195473, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195473
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195473

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACACCCACTGAGCTGG 24
|||||
Db 305 AAAAACACACCCACAGGTTGG 327

RESULT 13
US-09-949-016-16206/c
; Sequence 16206, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16206
; LENGTH: 54246
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16206

Query Match 72.8%; Score 18.2; DB 3; Length 54246;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACACCCACTGAGCTG 23
|||||

Best Local Similarity 87.0%; Pred. No. 1.1e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 20; Conservative 0;

Db 30983 AAAAACACACCCACAGCTG 30961

RESULT 14
US-09-949-016-13769
; Sequence 13769, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13769
; LENGTH: 58844
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13769

Query Match 72.8%; Score 18.2; DB 3; Length 58844;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACACCCACTGAGCTGG 24
|||||
Db 49085 AAAAACACCCAGCAGAGCTGG 49107

RESULT 15
US-09-949-016-17292/c
; Sequence 17292, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17292
; LENGTH: 112114
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17292

Query Match 72.8%; Score 18.2; DB 3; Length 112114;
Best Local Similarity 87.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACACCCACTGAGCTGG 24
|||||
Db 57932 AAAAACACACCCACAGGTTGG 57910

RESULT 16
US-09-949-016-11894/c
```

```
; Sequence 11894, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11894
; LENGTH: 154605
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(455726)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11894

Query Match 72.8%; Score 18.2; DB 3; Length 154605;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACCACTGAGCTGG 24
   |||||
Db 57930 AAAAAACACCACTGAGCTGG 57908

RESULT 17
US-09-949-016-14157/c
; Sequence 14157, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14157
; LENGTH: 455726
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(455726)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14157

Query Match 72.8%; Score 18.2; DB 3; Length 455726;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTG 23
   |||||
Db 238030 AAAAAACACCACTGAGCTG 238008

RESULT 18
US-09-949-016-11940/c
; Sequence 11940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11940
; LENGTH: 481115
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(481115)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11940

Query Match 72.8%; Score 18.2; DB 3; Length 481115;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTG 23
   |||||
Db 215419 AAAAAACACCACTGAGCTG 215397

RESULT 19
US-08-858-207A-71/c
; Sequence 71, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimm, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
```



```
;
; TELEX:
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1448 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; US-08-858-207A-71
;
; Query Match          71.2%; Score 17.8; DB 3; Length 1448;
; Best Local Similarity 90.5%; Pred. No. 1.8e+02;
; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
; QY 3 AAAACAACCACTGAGCTG 23
;      |||||
; DB 1225 AAAACAACCACTGAGCTG 1205
;
; RESULT 20
; US-09-489-039A-5144
; Sequence 5144, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5144
;   LENGTH: 1509
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-5144
;
; Query Match          71.2%; Score 17.8; DB 3; Length 1509;
; Best Local Similarity 90.5%; Pred. No. 1.8e+02;
; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
; QY 3 AAAACAACCACTGAGCTG 23
;      |||||
; DB 379 AAAACAACCACTGAGCTG 399
;
; RESULT 21
; US-08-865-311-1
; Sequence 1, Application US/08865311
; Patent No. 5919664
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; TITLE OF INVENTION: No. 5919664el Compounds
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/865,311
; FILING DATE: 29-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
```

```
;
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Falk, Stephen T
;   REGISTRATION NUMBER: 36,795
;   REFERENCE/DOCKET NUMBER: GM10004
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2488
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1542 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; US-08-865-311-1
;
; Query Match          71.2%; Score 17.8; DB 2; Length 1542;
; Best Local Similarity 90.5%; Pred. No. 1.8e+02;
; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
; QY 3 AAAACAACCACTGAGCTG 23
;      |||||
; DB 67 AAAACAACCACTGAGCTG 87
;
; RESULT 22
; US-09-315-720-1
; Sequence 1, Application US/09315720
; Patent No. 6372487
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; TITLE OF INVENTION: No. 6372487el Compounds
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stree
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,720
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/865,311
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Falk, Stephen T
;   REGISTRATION NUMBER: 36,795
;   REFERENCE/DOCKET NUMBER: GM10004
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2488
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1542 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; US-09-315-720-1
;
; Query Match          71.2%; Score 17.8; DB 3; Length 1542;
; Best Local Similarity 90.5%; Pred. No. 1.8e+02;
; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 AAAACAACCACTGAGCTG 23
Db 67 AAAACAACCACTGAGCAG 87

RESULT 23
US-09-583-110-1228
; Sequence 1228, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1228
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-1228

Query Match 71.2%; Score 17.8; DB 3; Length 1545;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACAACCACTGAGCTG 23
Db 67 AAAACAACCACTGAGCAG 87

RESULT 24
US-09-107-433-1292
; Sequence 1292, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; TITLE OF INVENTION: SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1292:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1566
; SEQUENCE DESCRIPTION: SEQ ID NO: 1292:
US-09-107-433-1292

Query Match 71.2%; Score 17.8; DB 3; Length 1566;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACAACCACTGAGCTG 23
Db 88 AAAACAACCACTGAGCAG 108

RESULT 25
US-08-961-527-7/c
; Sequence 7, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-7

Query Match 71.2%; Score 17.8; DB 3; Length 19702;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 3 AAAAAACACCACTGAGCTG 23  
|||||  
Db 2644 AAAAAACCACTGAGCAG 2624

RESULT 26  
US-09-949-016-16822/c  
; Sequence 16822, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16822  
; LENGTH: 73818  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(73818)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-16822

Query Match 71.2%; Score 17.8; DB 3; Length 73818;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGC 21  
|||||  
Db 30080 AAAAAACCACTGAGC 30060

RESULT 27  
US-09-949-016-24377/c  
; Sequence 24377, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 24377  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-24377

Query Match 70.4%; Score 17.6; DB 3; Length 601;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTG 24  
|||||  
Db 150 AAAAAACCACTGAGCTG 127

RESULT 28  
US-09-949-016-134073  
; Sequence 134073, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 134073  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-134073

Query Match 70.4%; Score 17.6; DB 3; Length 601;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACACCACTGAGCTG 25  
|||||  
Db 358 AAAAAACCACTGAGCTG 381

RESULT 29  
US-09-949-016-199548/c  
; Sequence 199548, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 199548  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-199548

Query Match 70.4%; Score 17.6; DB 3; Length 601;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTG 24  
|||||  
Db 150 AAAAAACCACTGAGCTG 127

```
RESULT 30
US-09-949-016-15944/c
; Sequence 15944, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15944
; LENGTH: 93532
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15944

Query Match          70.4%; Score 17.6; DB 3; Length 93532;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGG 24
Db 40655 AAAAAACACCAATCAGTGAGCTTG 40632

RESULT 31
US-09-949-016-12173
; Sequence 12173, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12173
; LENGTH: 221958
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12173

Query Match          70.4%; Score 17.6; DB 3; Length 221958;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACACCACTGAGCTGG 25
Db 125285 AAAAAACACCACTGAGCTGG 125308

RESULT 32
US-09-949-016-15498
; Sequence 15498, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15498
; LENGTH: 221966
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15498

Query Match          70.4%; Score 17.6; DB 3; Length 221966;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACACCACTGAGCTGG 25
Db 125285 AAAAAACACCACTGAGCTGG 125308

RESULT 33
US-09-949-016-17371/c
; Sequence 17371, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17371
; LENGTH: 373182
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(373182)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17371

Query Match          70.4%; Score 17.6; DB 3; Length 373182;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGG 24
Db 67405 AAACAACAAAGAACTGAGCTGG 67382

RESULT 34
US-09-949-016-12062/c
; Sequence 12062, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```



```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15954
; LENGTH: 54576
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(54576)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15954

Query Match          69.6%; Score 17.4; DB 3; Length 54576;
Best Local Similarity 94.7%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGA 19
Db 54461 AAAAAACACCACTGA 54479

RESULT 39
US-09-949-016-15955
; Sequence 15955, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15955
; LENGTH: 54576
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(54576)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15955

Query Match          69.6%; Score 17.4; DB 3; Length 54576;
Best Local Similarity 94.7%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGA 19
Db 54461 AAAAAACACCACTGA 54479

RESULT 40
US-09-949-016-13734/c
; Sequence 13734, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13734
; LENGTH: 131254
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(131254)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13734

Query Match          69.6%; Score 17.4; DB 3; Length 131254;
Best Local Similarity 94.7%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAACAACCACTGAGCT 22
Db 127059 AAAGAACCACCACTGAGCT 127041

RESULT 41
US-09-949-016-22643
; Sequence 22643, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22643
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(54576)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-22643

Query Match          68.8%; Score 17.2; DB 3; Length 601;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACCACTGAGCTG 23
Db 509 AAAAAACCACTGATCTG 530

RESULT 42
US-09-949-016-36627
; Sequence 36627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 36627  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-36627

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACAACCACTGAGCTGGG 25  
|||||  
Db 508 AAACGACCTCCACTGAGCTGGG 529

## RESULT 43

US-09-949-016-44701  
; Sequence 44701, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 44701  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-44701

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACAACCACTGAGCTGGG 25  
|||||  
Db 508 AAACGACCTCCACTGAGCTGGG 529

## RESULT 44

US-09-949-016-59992  
; Sequence 59992, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 59992  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-59992

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAACCACTGAGCT 22  
|||||  
Db 293 AAAAAAACCACTGAGCT 314

## RESULT 45

US-09-949-016-172749  
; Sequence 172749, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 172749  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-172749

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAACCACTGAGCTG 23  
|||||  
Db 509 ATAACAACCACTGATCTG 530

## RESULT 46

US-09-949-016-184440/c  
; Sequence 184440, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 184440  
; LENGTH: 601  
; TYPE: DNA





```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1278
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1278

Query Match      68.8%; Score 17.2; DB 3; Length 2103;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 AAACACCCACTGAGCTGGG 25
Db      2040 AAACGACCTCCACTGAACCTGGG 2061
```

Search completed: February 3, 2006, 16:31:57  
Job time : 89.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-3

Perfect score: 25

Sequence: 1 aaaaacaccaccactgagctggg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA\_Main:

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US11\_PUBCOMB.seq:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	US-10-719-900-3
2	23.4	93.6	25	8	US-10-719-900-4
3	20.2	80.8	512	5	US-10-027-632-48258
4	20.2	80.8	512	6	US-10-027-632-48258
5	20.2	80.8	568	4	US-09-925-065A-201949
6	20.2	80.8	620	5	US-10-027-632-75571
7	20.2	80.8	620	5	US-10-027-632-76226
8	20.2	80.8	620	6	US-10-027-632-76226
9	20.2	80.8	620	6	US-10-027-632-76226
10	19.8	79.2	556	4	US-09-925-065A-148340
11	19.2	76.8	201	8	US-10-719-993-34439
12	19.2	76.8	201	8	US-10-719-993-34440
13	19.2	76.8	503	3	US-09-918-995-31484
14	19.2	76.8	546025	8	US-10-719-993-31484
15	18.8	75.2	42061	7	US-10-322-281-770
16	18.8	75.2	96898	6	US-10-417-476-3
17	18.8	75.2	3673778	6	US-10-312-841-2
18	18.6	74.4	90	9	US-10-467-851-264
19	18.6	74.4	276	8	US-10-425-115-58262
20	18.6	74.4	333	7	US-10-242-535A-8589
21	18.6	74.4	333	7	US-10-085-783A-8589
22	18.6	74.4	540	7	US-10-276-774-680
23	18.6	74.4	709	4	US-09-925-065A-934362

24	18.6	74.4	755	5	US-10-027-632-129384	Sequence 129384,
25	18.6	74.4	755	6	US-10-027-632-129384	Sequence 129384,
c 26	18.6	74.4	777	7	US-10-343-650A-221	Sequence 221, App
c 27	18.6	74.4	882	5	US-10-027-633-129760	Sequence 129760,
c 28	18.6	74.4	882	5	US-10-027-633-129761	Sequence 129761,
c 29	18.6	74.4	882	5	US-10-027-633-129762	Sequence 129762,
c 30	18.6	74.4	882	5	US-10-027-633-129762	Sequence 129762,
c 31	18.6	74.4	882	6	US-10-027-633-129760	Sequence 129760,
c 32	18.6	74.4	882	6	US-10-027-633-129761	Sequence 129761,
c 33	18.6	74.4	882	6	US-10-027-633-129762	Sequence 129762,
c 34	18.6	74.4	882	6	US-10-027-633-129762	Sequence 129762,
c 35	18.6	74.4	962	4	US-09-925-065A-82610	Sequence 82610, A
c 36	18.6	74.4	1065	6	US-10-220-382-32	Sequence 32, Appl
c 37	18.6	74.4	1329	3	US-09-965-422-1	Sequence 1, Appl
c 38	18.6	74.4	1450	5	US-10-012-600B-103	Sequence 103, App
c 39	18.6	74.4	2028	3	US-09-981-566A-43	Sequence 43, Appl
c 40	18.6	74.4	2464	3	US-09-764-891-6728	Sequence 6728, Ap
c 41	18.6	74.4	2464	3	US-09-764-891-6729	Sequence 6729, Ap
c 42	18.6	74.4	3029	8	US-10-723-860-7237	Sequence 7237, Ap
c 43	18.6	74.4	3154	6	US-10-172-118-1308	Sequence 1308, Ap
c 44	18.6	74.4	3154	7	US-10-342-887-1308	Sequence 1308, Ap
c 45	18.6	74.4	22929	6	US-10-017-161-591	Sequence 591, App
c 46	18.6	74.4	122923	5	US-10-087-192-268	Sequence 268, App
c 47	18.6	74.4	128993	8	US-10-484-577-681	Sequence 681, App
c 48	18.6	74.4	177380	8	US-10-484-577-683	Sequence 683, App
c 49	18.4	73.6	607	5	US-10-027-632-279949	Sequence 279949,
c 50	18.4	73.6	607	5	US-10-027-632-279950	Sequence 279950,
c 51	18.4	73.6	607	6	US-10-027-633-279949	Sequence 279949,
c 52	18.4	73.6	607	6	US-10-027-632-279950	Sequence 279950,
c 53	18.4	73.6	2097	7	US-10-282-122A-40836	Sequence 40836, A
c 54	18.4	73.6	47756	7	US-10-322-696-148	Sequence 148, App
c 55	18.2	72.8	47756	8	US-10-719-993-28266	Sequence 28266, A
c 56	18.2	72.8	485	7	US-10-424-599-60162	Sequence 60162, A
c 57	18.2	72.8	553	4	US-09-925-065A-832430	Sequence 832430,
c 58	18.2	72.8	592	4	US-09-925-065A-448402	Sequence 448402,
c 59	18.2	72.8	592	4	US-09-925-065A-448403	Sequence 448403,
c 60	18.2	72.8	598	3	US-09-864-761-7372	Sequence 7372, Ap
c 61	18.2	72.8	599	9	US-10-972-079-71589	Sequence 71589, A
c 62	18.2	72.8	600	9	US-10-972-079-71590	Sequence 71590, A
c 63	18.2	72.8	600	9	US-10-972-079-71591	Sequence 71591, A
c 64	18.2	72.8	600	9	US-10-972-079-71592	Sequence 71592, A
c 65	18.2	72.8	600	9	US-10-972-079-92735	Sequence 92735, A
c 66	18.2	72.8	616	4	US-09-925-065A-952141	Sequence 952141,
c 67	18.2	72.8	652	5	US-10-027-632-242984	Sequence 242984,
c 68	18.2	72.8	652	6	US-10-027-632-242984	Sequence 242984,
c 69	18.2	72.8	1165	7	US-10-424-599-106254	Sequence 106254,
c 70	18.2	72.8	1383	4	US-09-925-065A-88703	Sequence 88703, A
c 71	18.2	72.8	4624	10	US-11-097-143-30992	Sequence 30992, A
c 72	18.2	72.8	326014	3	US-09-731-231A-3	Sequence 3, Appli
c 73	18.2	72.8	326014	7	US-10-751-985-3	Sequence 3, Appli
c 74	18.2	72.8	493999	8	US-10-719-993-6787	Sequence 6787, Ap
c 75	17.8	71.2	108	3	US-09-864-761-28125	Sequence 28125, A
c 76	17.8	71.2	352	3	US-09-764-864-655	Sequence 659, App
c 77	17.8	71.2	352	3	US-09-764-891-419	Sequence 419, App
c 78	17.8	71.2	352	6	US-10-091-414-26	Sequence 26, Appl
c 79	17.8	71.2	360	3	US-09-864-761-11537	Sequence 11537, A
c 80	17.8	71.2	427	7	US-10-437-963-31341	Sequence 31341, A
c 81	17.8	71.2	435	9	US-10-450-763-3205	Sequence 3205, Ap
c 82	17.8	71.2	504	6	US-10-029-386-6909	Sequence 6909, Ap
c 83	17.8	71.2	606	6	US-10-029-386-22987	Sequence 22987, A
c 84	17.8	71.2	1485	7	US-10-282-122A-23394	Sequence 23394, A
c 85	17.8	71.2	1542	8	US-10-472-928-725	Sequence 725, App
c 86	17.8	71.2	1545	7	US-10-282-122A-37633	Sequence 37633, A
c 87	17.8	71.2	1566	9	US-10-617-320-1292	Sequence 1292, Ap
c 88	17.8	71.2	1776	9	US-10-450-763-29300	Sequence 29300, A
c 89	17.8	71.2	1895	3	US-09-764-864-230	Sequence 230, App
c 90	17.8	71.2	2056	6	US-10-172-118-1662	Sequence 1662, Ap
c 91	17.8	71.2	2056	7	US-10-342-887-1662	Sequence 1662, Ap
c 92	17.8	71.2	5664	9	US-10-450-763-3207	Sequence 3207, Ap
c 93	17.8	71.2	19702	2	US-08-961-527-7	Sequence 7, Appli
c 94	17.8	71.2	19702	7	US-10-158-844-7	Sequence 7, Appli
c 95	17.8	71.2	26006	3	US-09-764-869-1963	Sequence 1963, Ap
c 96	17.8	71.2	26006	3	US-09-764-864-1638	Sequence 1638, Ap

```
c 97 17.8 71.2 26006 3 US-09-764-891-6285 Sequence 6285, Ap
c 98 17.8 71.2 26006 5 US-10-091-504-1963 Sequence 1963, Ap
c 99 17.8 71.2 26006 6 US-10-091-414-242 Sequence 242, App
c 100 17.8 71.2 26006 6 US-10-242-355-997 Sequence 997, App
c 101 17.8 71.2 26006 3 US-10-227-577-1963 Sequence 1963, Ap
c 102 17.8 71.2 26013 3 US-09-764-869-1961 Sequence 1961, Ap
c 103 17.8 71.2 26013 3 US-09-764-864-1636 Sequence 1636, Ap
c 104 17.8 71.2 26013 3 US-09-764-891-6283 Sequence 6283, Ap
c 105 17.8 71.2 26013 5 US-10-091-504-1961 Sequence 1961, Ap
c 106 17.8 71.2 26013 6 US-10-091-414-240 Sequence 240, App
c 107 17.8 71.2 26013 6 US-10-242-355-995 Sequence 995, App
c 108 17.8 71.2 26013 6 US-10-227-577-1961 Sequence 1961, Ap
c 109 17.8 71.2 26018 3 US-09-764-869-1962 Sequence 1962, Ap
c 110 17.8 71.2 26018 3 US-09-764-864-1637 Sequence 1637, Ap
c 111 17.8 71.2 26018 3 US-09-764-891-6284 Sequence 6284, Ap
c 112 17.8 71.2 26018 5 US-10-091-504-1962 Sequence 1962, Ap
c 113 17.8 71.2 26018 6 US-10-091-414-241 Sequence 241, App
c 114 17.8 71.2 26018 6 US-10-242-355-996 Sequence 996, App
c 115 17.8 71.2 26018 6 US-10-227-577-1962 Sequence 1962, Ap
c 116 17.8 71.2 26043 7 US-10-304-113-4 Sequence 4, Appli
c 117 17.8 71.2 230101 8 US-10-719-993-6829 Sequence 6829, Ap
c 118 17.8 71.2 2162598 8 US-10-472-928-4979 Sequence 4979, Ap
c 119 17.6 70.4 201 8 US-10-719-993-36574 Sequence 36574, A
c 120 17.6 70.4 265 3 US-09-560-863-148 Sequence 148, App
c 121 17.6 70.4 501 5 US-10-027-632-107363 Sequence 107363,
c 122 17.6 70.4 522 4 US-10-027-632-107363 Sequence 107363,
c 123 17.6 70.4 522 4 US-09-925-065A-64391 Sequence 64391, A
c 124 17.6 70.4 538 5 US-10-027-632-199273 Sequence 199273,
c 125 17.6 70.4 538 5 US-10-027-632-199273 Sequence 199273,
c 126 17.6 70.4 600 4 US-09-925-065A-116076 Sequence 116076,
c 127 17.6 70.4 600 4 US-09-925-065A-116077 Sequence 116077,
c 128 17.6 70.4 600 9 US-10-972-079-20258 Sequence 20258, A
c 129 17.6 70.4 600 9 US-10-972-079-20259 Sequence 20259, A
c 130 17.6 70.4 600 9 US-10-972-079-41751 Sequence 41751, A
c 131 17.6 70.4 600 9 US-10-972-079-53149 Sequence 53149, A
c 132 17.6 70.4 600 9 US-10-972-079-53150 Sequence 53150, A
c 133 17.6 70.4 617 4 US-09-925-065A-387867 Sequence 387867,
c 134 17.6 70.4 617 4 US-09-925-065A-387868 Sequence 387868,
c 135 17.6 70.4 617 4 US-09-925-065A-387869 Sequence 387869,
c 136 17.6 70.4 618 5 US-10-027-632-160009 Sequence 160009,
c 137 17.6 70.4 618 5 US-10-027-632-160009 Sequence 160009,
c 138 17.6 70.4 631 4 US-09-925-065A-687630 Sequence 687630,
c 139 17.6 70.4 631 4 US-09-925-065A-687631 Sequence 687631,
c 140 17.6 70.4 636 4 US-09-925-065A-359335 Sequence 359335,
c 141 17.6 70.4 764 3 US-09-770-445-941 Sequence 941, App
c 142 17.6 70.4 782 9 US-10-779-543-3972 Sequence 3972, Ap
c 143 17.6 70.4 1323 8 US-10-474-793-657 Sequence 657, App
c 144 17.6 70.4 1337 4 US-09-925-065A-704583 Sequence 704583,
c 145 17.6 70.4 1337 4 US-09-925-065A-704584 Sequence 704584,
c 146 17.6 70.4 1337 4 US-09-925-065A-704585 Sequence 704585,
c 147 17.6 70.4 1337 4 US-09-925-065A-704586 Sequence 704586,
c 148 17.6 70.4 3399 7 US-10-424-593-113687 Sequence 113687,
c 149 17.6 70.4 7386 7 US-10-114-270-193 Sequence 193, App
c 150 17.6 70.4 7386 7 US-10-114-270-193 Sequence 193, App

; SEQ ID NO 3
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-3

Query Match 100.0%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAACCACTGAGCTGGG 25
|||||
Db 1 AAAAAAACCACTGAGCTGGG 25
|||||

RESULT 2
US-10-719-900-4
; Sequence 4, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; PRIOR FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 4
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-4

Query Match 93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.86;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAACCACTGAGCTGGG 25
|||||
Db 1 AAAAAAACCACTGAGCTGGG 25
|||||

RESULT 3
US-10-027-632-48258/c
; Sequence 48258, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48258
; LENGTH: 512
```

ALIGNMENTS

```
US-10-719-900-3
; Sequence 3, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; PRIOR FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

```
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48258

Query Match      80.8%; Score 20.2; DB 5; Length 512;
Best Local Similarity 88.0%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 146 AAAAAACACCACTGATCTAGG 122

RESULT 4
US-10-027-632-48258/c
; Sequence 48258, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48258
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48258

Query Match      80.8%; Score 20.2; DB 6; Length 512;
Best Local Similarity 88.0%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 146 AAAAAACACCACTGATCTAGG 122

RESULT 5
US-09-925-065A-201949/c
; Sequence 201949, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766

; TYPE: DNA
; ORGANISM: Human
US-10-027-632-75571/c
; Sequence 75571, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75571
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(620)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-75571

Query Match      80.8%; Score 20.2; DB 5; Length 620;
Best Local Similarity 88.0%; Pred. No. 36;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
   ||||| ||||| ||||| ||||| |||||
Db 146 AAAAAACACCACTGATCTAGG 122

RESULT 7
US-10-027-632-76226/c
; Sequence 76226, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```





```

; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 148340
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-148340

```

Query Match 79.2%; Score 19.8; DB 4; Length 556;  
Best Local Similarity 84.0%; Pred. No. 54;  
Matches 21: Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	1	AAAAACAACCA	25
		: : : : : :	
pb	286	AAAAACAACCA	262
		: : : : : :	

```

RESULT 11
US-10-719-993-34439/c
; Sequence 34439, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 34439
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-34439

```

Query Match	76.8%	Score 19.2;	DB 8;	Length 201;
Best Local Similarity	87.5%;	Pred. No. 91;		
Matches	21.	Conservative	0:	Mismatches
			3:	Indels
			0:	Gaps
			0:	

Qy 2 AAAAAACACCACCACTGAGCTGGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 172 AAAAAATAACCAACACTGGGCTGGG 149

```

RESULT 12
US-10-719-993-34440/c
; Sequence 34440, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34440
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-34440

```

Query Match 76.8%; Score 19.2; DB 8; Length 201;  
Best Local Similarity 87.5%; Pred. No. 91;  
Matches 21: Conservative 0: Mismatches 3: Indels 0: Gaps 0;

**Qy**

2 AAAAAACACCACCTGAGCTGG 25  
|||||  
**Db**

128 AAAAAATACCAACTGGGTGG 105  
|||||

```

RESULT 13
US-09-918-995-31484/c
; Sequence 31484, Application US/09018995
; Publication NO. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hvsseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: Fastseq for Windows Version 3.0
; SEQ ID NO 31484
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(503)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31484

```

```
Query Match          76.8%; Score 19.2; DB 3; Length 503;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 2 AAAAAACCAACCACCTGAGCTGG 25  
||| ||||| ||||| |||||  
Db 280 AAATACAAACCAACCTGACCTGG 257

```

RESULT 14
US-10-719-993-6862/c
; Sequence 6862, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CAPGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 6862
; LENGTH: 546025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(546025)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tab
US-10-719-993-6862

```

Query Match	76.8%	Score 19.2;	DB 8;	Length 546025;
Best Local Similarity	87.5%	Pred. No. 2.2e+02;		
Matches 21: Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY 2 AAAAAACAACCACTGAGCTGG 25  
||||| ||||| ||||| ||||| |||||  
Db 246646 AAAAAACAACCACTGGCTGG 246623

## RESULT 15

US-10-322-281-770/c  
; Sequence 770, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 770  
; LENGTH: 42061  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(42061)  
; OTHER INFORMATION: n = A, T, C or G  
US-10-322-281-770

Query Match 75.2%; Score 18.8; DB 7; Length 42061;  
Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCT 22

Db 10151 AAAAAACACCACTGAGCT 10130

## RESULT 16

US-10-417-476-3  
; Sequence 3, Application US/10417476  
; Publication No. US20040002102A1  
; GENERAL INFORMATION:  
; APPLICANT: Litman, Gary W.  
; APPLICANT: Hawke, Jeffrey A.  
; APPLICANT: Yoder, Jeffrey A.  
; TITLE OF INVENTION: B1VM (Basic, Immunoglobulin-Like Variable Motif-Containing) Gene,  
; FILE REFERENCE: USF-103X  
; CURRENT APPLICATION NUMBER: US/10/417,476  
; CURRENT FILING DATE: 2003-04-16  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 96898  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (466)..(702)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 6  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (823)..(997)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 5  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2904)..(2985)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 4  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3164)..(3366)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 3  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6395)..(6570)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 2  
; FEATURE:  
; NAME/KEY: misc feature

; LOCATION: (8013)..(8312)  
; OTHER INFORMATION: Genomic fragment identified as part of CpG island (Genbank Z59762)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8059)..(8572)  
; OTHER INFORMATION: Inverse complement of MGC5302 Exon 1  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8223)..(8235)  
; OTHER INFORMATION: Inverse complement of MGC5302 translation initiation codon (ATG)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8614)..(9086)  
; OTHER INFORMATION: B1VM Exon A - untranslated  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9019)..(9033)  
; OTHER INFORMATION: B1VM Exon A alternative splice donor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9077)..(9091)  
; OTHER INFORMATION: B1VM Exon A splice donor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (14862)..(14876)  
; OTHER INFORMATION: B1VM Exon B splice acceptor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (14872)..(14955)  
; OTHER INFORMATION: B1VM Exon B - untranslated  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (14946)..(14960)  
; OTHER INFORMATION: B1VM Exon B splice donor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (16309)..(16309)  
; OTHER INFORMATION: n = a, c, g, or t.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (16701)..(16715)  
; OTHER INFORMATION: B1VM Exon 1 splice acceptor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (16711)..(17310)  
; OTHER INFORMATION: B1VM Exon 1  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (16833)..(16835)  
; OTHER INFORMATION: B1VM translation initiation codon (ATG)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (17301)..(17315)  
; OTHER INFORMATION: B1VM Exon 1 splice donor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (25983)..(25997)  
; OTHER INFORMATION: B1VM Exon 2 splice acceptor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (25993)..(26119)  
; OTHER INFORMATION: B1VM Exon 2  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (26110)..(26124)  
; OTHER INFORMATION: B1VM Exon 2 splice donor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (30592)..(30606)  
; OTHER INFORMATION: B1VM Exon 3 splice acceptor site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (30602)..(30697)

```
OTHER INFORMATION: BIVM Exon 3
FEATURE:
NAME/KEY: misc feature
LOCATION: (30688)..(30702)
OTHER INFORMATION: BIVM Exon 3 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (31298)..(31312)
OTHER INFORMATION: BIVM Exon 4 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (31308)..(31412)
OTHER INFORMATION: BIVM Exon 4
FEATURE:
NAME/KEY: misc feature
LOCATION: (31403)..(31417)
OTHER INFORMATION: BIVM Exon 4 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (31620)..(31634)
OTHER INFORMATION: BIVM Exon 5 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (31630)..(31724)
OTHER INFORMATION: BIVM Exon 5
FEATURE:
NAME/KEY: misc feature
LOCATION: (31715)..(31729)
OTHER INFORMATION: BIVM Exon 5 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (41120)..(41134)
OTHER INFORMATION: BIVM Exon 6 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (41130)..(41262)
OTHER INFORMATION: BIVM Exon 6
FEATURE:
NAME/KEY: misc feature
LOCATION: (41253)..(41267)
OTHER INFORMATION: BIVM Exon 6 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (44021)..(44035)
OTHER INFORMATION: BIVM Exon 7 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (44031)..(44117)
OTHER INFORMATION: BIVM Exon 7
FEATURE:
NAME/KEY: misc feature
LOCATION: (44108)..(44122)
OTHER INFORMATION: BIVM Exon 7 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (48198)..(48212)
OTHER INFORMATION: BIVM Exon 8 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (48208)..(48304)
OTHER INFORMATION: BIVM Exon 8
FEATURE:
NAME/KEY: misc feature
LOCATION: (48295)..(48309)
OTHER INFORMATION: BIVM Exon 8 splice donor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (49127)..(49141)
OTHER INFORMATION: BIVM Exon 9 splice acceptor site
FEATURE:
NAME/KEY: misc feature
LOCATION: (49137)..(51096)
OTHER INFORMATION: BIVM Exon 9
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (49428)..(49430)
OTHER INFORMATION: BIVM translation termination codon (TGA)
FEATURE:
NAME/KEY: misc feature
LOCATION: (50039)..(50248)
OTHER INFORMATION: Alu sequence
FEATURE:
NAME/KEY: misc feature
LOCATION: (55216)..(55975)
OTHER INFORMATION: ERCC5 Exon 1
FEATURE:
NAME/KEY: misc feature
```

Query Match 75.2%; Score 18.8; DB 6; Length 96898;  
Best Local Similarity 90.9%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCT 22  
|||||  
DB 5945 AAAAAACACCACTGAGCT 5966  
|||||

## RESULT 17

US-10-312-841-2/c  
Sequence 2, Application US/10312841  
Publication No. US20030186277A1  
GENERAL INFORMATION:  
APPLICANT: Epigenomics AG  
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHA  
FILE REFERENCE: E01/1208/WO  
CURRENT APPLICATION NUMBER: US/10/312,841  
CURRENT FILING DATE: 2002-12-30  
NUMBER OF SEQ ID NOS: 2  
SEQ ID NO 2  
LENGTH: 3673778  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (379615)  
US-10-312-841-2

Query Match 75.2%; Score 18.8; DB 6; Length 3673778;  
Best Local Similarity 90.9%; Pred. No. 3.8e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCT 22  
|||||  
DB 844278 AAAAAACACCACTATCT 844257  
|||||

## RESULT 18

US-10-467-851-264/c  
Sequence 264, Application US/10467851  
Publication No. US20050221303A1  
GENERAL INFORMATION:  
APPLICANT: TELERMAN, Adam  
APPLICANT: AMSON, Robert  
APPLICANT: TUIJNDER, Marcel,  
APPLICANT: SUSINI, Laurent  
TITLE OF INVENTION: SEQUENCES INVOLVED IN PHENOMENA OF TUMOUR SUPPRESSION,  
TITLE OF INVENTION: TUMOUR REVERSION, APOPTOSIS AND/OR VIRUS RESISTANCE  
TITLE OF INVENTION: AND THEIR USE AS MEDICINES  
FILE REFERENCE: 11416-014-999  
CURRENT APPLICATION NUMBER: US/10/467,851  
CURRENT FILING DATE: 2003-08-13  
PRIOR APPLICATION NUMBER: PCT/FR 02/00 543  
PRIOR FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: FR 01/01 925

```
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 1020
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 264
; LENGTH: 90
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-10-467-851-264

Query Match      74.4%; Score 18.6; DB 9; Length 90;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACCTGAGCTGGG 25
   |||||
Db 43 AAAAAACACCCACCTGAGCTGGG 19
   |||||

RESULT 19
US-10-425-115-58262/c
; Sequence 58262, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 58262
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(276)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: MRT4577_153111C.1
US-10-425-115-58262

Query Match      74.4%; Score 18.6; DB 8; Length 276;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACCTGAGCTGGG 25
   |||||
Db 129 AAAAAACAGGACCACTGAGCTGTG 105
   |||||

RESULT 20
US-10-242-535A-8589
; Sequence 8589, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
```

```
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8589
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (315)..(315)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-8589

Query Match      74.4%; Score 18.6; DB 7; Length 333;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACCTGAGCTGGG 25
   |||||
Db 279 AAAAAACACCCACCTGAGCTGGG 303
   |||||

RESULT 21
US-10-085-783A-8589
; Sequence 8589, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8589
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (315)..(315)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-8589

Query Match      74.4%; Score 18.6; DB 7; Length 333;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACCTGAGCTGGG 25
   |||||
Db 279 AAAAAACACCCACCTGAGCTGGG 303
   |||||

RESULT 22
US-10-278-774-680
; Sequence 680, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 129384
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-129384

Query Match          74.4%; Score 18.6; DB 5; Length 755;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCTGGG 25
   |||||
Db 588 AAAAAACAACCACTGAGCTGGG 612

RESULT 25
US-10-027-632-129384
; Sequence 129384, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMERIZATION: Polymorphisms in the Human Genome
; FILE REFERENCES: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 129384
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-129384

Query Match          74.4%; Score 18.6; DB 6; Length 755;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCTGGG 25
   |||||
Db 588 AAAAAACAACCACTGAGCTGGG 612

```

## RESULT 26

US-10-343-650A-221/c  
; Sequence 221, Application US/10343650A  
; Publication No. US20040067499A1  
; GENERAL INFORMATION:  
; APPLICANT: HAGA, TATSUYA  
; TITLE OF INVENTION: NOVEL G-PROTEIN COUPLED RECEPTOR  
; FILE REFERENCE: 31671-186347  
; CURRENT APPLICATION NUMBER: US/10/343,650A  
; PRIOR FILING DATE: 2003-07-21  
; PRIOR APPLICATION NUMBER: JP 2000/237818  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: JP 2001/34434  
; PRIOR FILING DATE: 2001-02-13  
; NUMBER OF SEQ ID NOS: 694  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 221  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(777)  
US-10-343-650A-221

Query Match 74.4%; Score 18.6; DB 7; Length 777;  
Best Local Similarity 84.0%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 544 AAAAAACACCACTGTCGGTG 520

## RESULT 27

US-10-027-632-129760/c  
; Sequence 129760, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 129760  
; LENGTH: 882  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-129760

Query Match 74.4%; Score 18.6; DB 5; Length 882;  
Best Local Similarity 84.0%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 594 AAAAAACACCACTGACTGGG 570

## RESULT 28

US-10-027-632-129761/c  
; Sequence 129761, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 129761  
; LENGTH: 882  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-129761

Query Match 74.4%; Score 18.6; DB 5; Length 882;  
Best Local Similarity 84.0%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25

|||||

Db 594 AAAAAACACCACTGACTGGG 570

## RESULT 29

US-10-027-632-129762/c  
; Sequence 129762, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720





```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129762
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-129762

Query Match 74.4%; Score 18.6; DB 6; Length 882;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 594 AAAAAACACCACTGAGCTGGG 570

RESULT 34
US-10-027-632-159900/c
; Sequence 159900, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159900
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-159900

Query Match 74.4%; Score 18.6; DB 6; Length 882;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 594 AAAAAACACCACTGAGCTGGG 570

RESULT 35
US-09-925-065A-82610/c
; Sequence 82610, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82610
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-82610

Query Match 74.4%; Score 18.6; DB 4; Length 962;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 549 AAAAAACACCACTGAGCTGGG 525

RESULT 36
US-10-220-382-32/c
; Sequence 32, Application US/10220382
; Publication No. US2003011911A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAU, Preeti
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: YAO, Monique G.
; APPLICANT: SHIH, Leo L.
; APPLICANT: TRIBOULEY, Catherine
; APPLICANT: LIJ, Dyung Aina M.
; APPLICANT: YUE, Henry
; APPLICANT: KHAN, Farrah A.
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: YANG, Junming
; APPLICANT: HARLAND, Lee
; APPLICANT: WALSH, Roderick T.
; APPLICANT: LO, Terence P.
; APPLICANT: BOROWSKY, Mark L.
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PI-0044 PCT
; CURRENT APPLICATION NUMBER: US/10/220,382
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,854; 60/188,384; 60/190,453; 60/190,730
; PRIOR FILING DATE: 2000-03-03; 2000-03-10; 2000-03-17; 2000-03-20

```

```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129762
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-129762

Query Match 74.4%; Score 18.6; DB 6; Length 882;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 594 AAAAAACACCACTGAGCTGGG 570

RESULT 34
US-10-027-632-159900/c
; Sequence 159900, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159900
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-159900

Query Match 74.4%; Score 18.6; DB 6; Length 882;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 594 AAAAAACACCACTGAGCTGGG 570

RESULT 35
US-09-925-065A-82610/c
; Sequence 82610, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82610
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-82610

Query Match 74.4%; Score 18.6; DB 4; Length 962;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 549 AAAAAACACCACTGAGCTGGG 525

RESULT 36
US-10-220-382-32/c
; Sequence 32, Application US/10220382
; Publication No. US2003011911A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAU, Preeti
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: YAO, Monique G.
; APPLICANT: SHIH, Leo L.
; APPLICANT: TRIBOULEY, Catherine
; APPLICANT: LIJ, Dyung Aina M.
; APPLICANT: YUE, Henry
; APPLICANT: KHAN, Farrah A.
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: YANG, Junming
; APPLICANT: HARLAND, Lee
; APPLICANT: WALSH, Roderick T.
; APPLICANT: LO, Terence P.
; APPLICANT: BOROWSKY, Mark L.
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: PI-0044 PCT
; CURRENT APPLICATION NUMBER: US/10/220,382
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,854; 60/188,384; 60/190,453; 60/190,730
; PRIOR FILING DATE: 2000-03-03; 2000-03-10; 2000-03-17; 2000-03-20

```



```
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/245,484
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/255,017
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/263,216
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/268,225
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 209
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-981-566A-43

Query Match      74.4%; Score 18.6; DB 3; Length 2028;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 1216 AAAAAACCACTGCTGGGTG 1192

RESULT 40
US-09-764-891-6728/c
; Sequence 6728, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6728
; LENGTH: 2464
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6728

Query Match      74.4%; Score 18.6; DB 3; Length 2464;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 714 AAAAAACAAACCACTGAACTGGG 690

RESULT 41
US-09-764-891-6729/c
; Sequence 6729, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6729
; LENGTH: 2464
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6729

Query Match      74.4%; Score 18.6; DB 3; Length 2464;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 714 AAAAAACAAACCACTGAACTGGG 690

RESULT 42
US-10-723-860-7237
; Sequence 7237, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NFUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7237
; LENGTH: 3029
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (705)..(722)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1012)..(1026)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7237

Query Match      74.4%; Score 18.6; DB 8; Length 3029;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTGAGCTGGG 25
Db 2946 AAAAAACCAACCACTGAAATGTG 2970

RESULT 43
US-10-172-118-1308
; Sequence 1308, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1308
; LENGTH: 3154
; TYPE: DNA
; ORGANISM: Homo sapiens
```



```
Db 12267 AAAAAAAAAACCAACAGAGCTAGG 12243
||||| ||||| ||||| || ||||| ||
RESULT 47
US-10-484-577-681
; Sequence 681, Application US/10484577
; Publication No. US20050032724A1
; GENERAL INFORMATION:
; APPLICANT: EPIDAUROS Biotechnologie Aktiengesellschaft
; TITLE OF INVENTION: Means and methods for improved treatment of cancer based on UGT1A
; FILE REFERENCE: F2285PCT-1
; CURRENT APPLICATION NUMBER: US/10/484,577
; CURRENT FILING DATE: 2004-01-22
; PRIOR APPLICATION NUMBER: PCT/EP 02/08220
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: EP 01 11 7608.8
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: EP 02011710.7
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 683
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 681
; LENGTH: 128993
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(128993
; LOCATION: )
; OTHER INFORMATION: n=a, c, g or t
US-10-484-577-681
Query Match 74.4%; Score 18.6; DB 8; Length 128993;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCAACCTGAGCTGGG 25
||||| ||||| ||||| ||||| |||||
Db 121849 AAAAGACACCCCTCACTGAGCTGGG 121873
RESULT 48
US-10-484-577-683/c
; Sequence 683, Application US/10484577
; Publication No. US20050032724A1
; GENERAL INFORMATION:
; APPLICANT: EPIDAUROS Biotechnologie Aktiengesellschaft
; TITLE OF INVENTION: Means and methods for improved treatment of cancer based on UGT1A
; FILE REFERENCE: F2285PCT-1
; CURRENT APPLICATION NUMBER: US/10/484,577
; CURRENT FILING DATE: 2004-01-22
; PRIOR APPLICATION NUMBER: PCT/EP 02/08220
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: EP 01 11 7608.8
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: EP 02011710.7
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 683
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 683
; LENGTH: 177380
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-484-577-683
Query Match 74.4%; Score 18.6; DB 8; Length 177380;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCAACCTGAGCTGGG 25
||||| ||||| ||||| ||||| |||||
Db 153700 AAAAGACACCCCTCACTGAGCTGGG 153676
```

```
RESULT 49
US-10-027-632-279949/c
; Sequence 279949, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279949
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-279949
Query Match 73.6%; Score 18.4; DB 5; Length 607;
Best Local Similarity 95.0%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAACAACCAACCACTGAGCTG 23
||||| ||||| ||||| |||||
Db 492 AAACAACCAACCACTGAGCTG 473
RESULT 50
US-10-027-632-279950/c
; Sequence 279950, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 279950
; LENGTH: 607
```

Mon Feb 6 12:23:16 2006

; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-279950  
Query Match 73.6%; Score 18.4; DB 5; Length 607;  
Best Local Similarity 95.0%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 4 AACACACCACCACTGAGCTG 23  
Db 492 AACACACCACCACTGAGCTG 473

Search completed: February 3, 2006, 15:43:28  
Job time : 375.556 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 / Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-3

Perfect score: 25

Sequence: 1 aaaaacacaccactgagctggg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 150 summaries

Database : Published Applications NA New:

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	18.6	74.4	1792	7	US-10-750-185-37447
2	18.6	74.4	1792	7	US-10-750-185-37447
3	18.6	74.4	86950	7	US-10-750-185-37447
4	18.4	73.6	1806	7	US-10-750-185-45680
5	18.4	73.6	1806	7	US-10-750-185-45680
6	17.8	71.2	600	8	US-11-136-527-4239
7	17.8	71.2	5383	8	US-11-136-527-4239
8	17.6	70.4	201	7	US-10-995-561-13216
9	17.6	70.4	2817	8	US-11-136-527-4239
10	17.6	70.4	380749	7	US-10-995-561-13216
11	17.4	69.6	21	7	US-10-310-914A-172254
12	17.2	68.8	25	8	US-11-121-849-30028
13	17.2	68.8	1188	7	US-10-750-185-34090
14	17.2	68.8	1188	7	US-10-750-185-34090
15	17.2	68.8	1324	7	US-10-750-185-25618
16	17.2	68.8	1324	7	US-10-750-185-25618
17	17.2	68.8	1826	7	US-10-750-185-27616
18	17.2	68.8	1826	7	US-10-750-185-27616
19	17.0	68.0	201	7	US-10-995-561-13216
20	17.0	68.0	201	7	US-10-995-561-13216
21	17.0	68.0	201	7	US-10-995-561-13216
22	17.0	68.0	201	7	US-10-995-561-13216

```

96 16.2 64.8 193084 8 US-11-121-086-82
97 16.2 64.8 207835 8 US-11-121-086-39
98 16.2 64.8 207835 8 US-11-121-086-40
99 16.2 64.8 220895 7 US-10-775-169-88
C 100 16 64.0 64 7 US-10-310-914A-20171
C 101 16 64.0 109 7 US-10-310-914A-10751
C 102 16 64.0 201 7 US-10-995-561-27356
C 103 16 64.0 201 7 US-10-995-561-27358
C 104 16 64.0 201 7 US-10-995-561-41476
C 105 16 64.0 201 7 US-10-995-561-41478
C 106 16 64.0 201 7 US-10-995-561-79494
C 107 16 64.0 201 7 US-10-995-561-79495
C 108 16 64.0 201 7 US-10-995-561-79497
C 109 16 64.0 201 7 US-10-995-561-79498
C 110 16 64.0 201 7 US-10-995-561-79499
C 111 16 64.0 201 7 US-10-995-561-79500
C 112 16 64.0 201 7 US-10-995-561-79501
C 113 16 64.0 201 7 US-10-995-561-79502
C 114 16 64.0 201 7 US-10-995-561-79514
C 115 16 64.0 201 7 US-10-995-561-79515
C 116 16 64.0 588 8 US-11-052-554A-666
C 117 16 64.0 600 7 US-10-750-185-24
C 118 16 64.0 600 7 US-10-750-185-1493
C 119 16 64.0 600 7 US-10-750-623-24
C 120 16 64.0 600 7 US-10-750-623-1493
C 121 16 64.0 619 8 US-11-000-463-650
C 122 16 64.0 664 8 US-11-000-463-178
C 123 16 64.0 873 7 US-10-750-185-48965
C 124 16 64.0 873 7 US-10-750-623-48965
C 125 16 64.0 892 7 US-10-750-185-56516
C 126 16 64.0 892 7 US-10-750-623-56516
C 127 16 64.0 1041 7 US-10-750-185-39741
C 128 16 64.0 1041 7 US-10-750-623-39741
C 129 16 64.0 1042 7 US-10-750-185-62101
C 130 16 64.0 1042 7 US-10-750-623-62101
C 131 16 64.0 1148 8 US-11-099-691-24
C 132 16 64.0 1478 7 US-10-750-185-46236
C 133 16 64.0 1478 7 US-10-750-623-46236
C 134 16 64.0 1569 7 US-10-750-185-26134
C 135 16 64.0 1569 7 US-10-750-623-26134
C 136 16 64.0 1762 7 US-10-750-185-58016
C 137 16 64.0 1762 7 US-10-750-623-58016
C 138 16 64.0 1807 7 US-10-750-185-30577
C 139 16 64.0 1807 7 US-10-750-623-30577
C 140 16 64.0 2080 7 US-10-750-185-36859
C 141 16 64.0 2080 7 US-10-750-623-36859
C 142 16 64.0 2346 8 US-11-102-026A-8
C 143 16 64.0 2485 7 US-10-750-185-55627
C 144 16 64.0 2485 7 US-10-750-623-55627
C 145 16 64.0 2489 7 US-10-750-185-61906
C 146 16 64.0 2489 7 US-10-750-623-61906
C 147 16 64.0 2624 7 US-10-750-185-32378
C 148 16 64.0 2624 7 US-10-750-623-32378
C 149 16 64.0 2742 8 US-11-169-041-30
C 150 16 64.0 2743 7 US-10-821-234-267
```

ALIGNMENTS

```

RESULT 1
US-10-750-185-37447
; Sequence 37447, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

RESULT 3
US-10-750-185-37447
; Sequence 37447, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

RESULT 3
US-10-857-780-5
; Sequence 5, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELEND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28

RESULT 3
US-10-857-780-5
; Sequence 5, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELEND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
```

; PRIOR APPLICATION NUMBER: 10/723,681  
 ; PRIOR FILING DATE: 2003-11-25  
 ; PRIOR APPLICATION NUMBER: 60/490,234  
 ; PRIOR FILING DATE: 2003-07-24  
 ; PRIOR APPLICATION NUMBER: 60/525,239  
 ; PRIOR FILING DATE: 2003-11-25  
 ; NUMBER OF SEQ ID NOS: 4962  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 5  
 ; LENGTH: 86950  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (20282)..(20282)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (39985)..(39985)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (39989)..(39989)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (51996)..(51996)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (64732)..(64732)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (83916)..(83919)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (86196)..(86196)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; US-10-857-780-5

Query Match 74.4%; Score 18.6; DB 7; Length 86950;  
 Best Local Similarity 84.0%; Pred. No. 73;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCTGGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 42395 AAAAAAGATTACCTGGCTGGG 42419

RESULT 4  
 US-10-750-185-45680/c  
 ; Sequence 45680, Application US/10750185  
 ; Publication No. US20050260603A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MMI GENOMICS, INC.  
 ; APPLICANT: DENISE, Sue K.  
 ; APPLICANT: KERR, Richard  
 ; APPLICANT: ROSENFELD, David  
 ; APPLICANT: HOLM, Tom  
 ; APPLICANT: BATES, Stephen  
 ; APPLICANT: FANTIN, Dennis  
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
 ; FILE REFERENCE: MM1100-2  
 ; CURRENT APPLICATION NUMBER: US/10750,185  
 ; CURRENT FILING DATE: 2003-12-31  
 ; PRIOR APPLICATION NUMBER: US 60/437,482  
 ; PRIOR FILING DATE: 2002-12-31  
 ; NUMBER OF SEQ ID NOS: 64922  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 45680  
 ; LENGTH: 1806

; TYPE: DNA  
 ; ORGANISM: Bovine 19866880824813  
 ; US-10-750-185-45680

Query Match 73.6%; Score 18.4; DB 7; Length 1806;  
 Best Local Similarity 95.0%; Pred. No. 45;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAG 20  
 ||||| ||||| ||||| ||||| |||||  
 Db 192 AAAAAACAACCACTGAG 173

RESULT 5  
 US-10-750-623-45680/c  
 ; Sequence 45680, Application US/10750623  
 ; Publication No. US20050287531A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MMI GENOMICS, INC.  
 ; APPLICANT: DENISE, Sue K.  
 ; APPLICANT: KERR, Richard  
 ; APPLICANT: ROSENFELD, David  
 ; APPLICANT: HOLM, Tom  
 ; APPLICANT: BATES, Stephen  
 ; APPLICANT: FANTIN, Dennis  
 ; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
 ; FILE REFERENCE: MM1100-1  
 ; CURRENT APPLICATION NUMBER: US/10750,623  
 ; CURRENT FILING DATE: 2003-12-31  
 ; PRIOR APPLICATION NUMBER: US 60/437,482  
 ; PRIOR FILING DATE: 2002-12-31  
 ; NUMBER OF SEQ ID NOS: 64922  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 45680  
 ; LENGTH: 1806  
 ; TYPE: DNA  
 ; ORGANISM: Bovine 19866880824813  
 ; US-10-750-623-45680

Query Match 73.6%; Score 18.4; DB 7; Length 1806;  
 Best Local Similarity 95.0%; Pred. No. 45;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAG 20  
 ||||| ||||| ||||| ||||| |||||  
 Db 192 AAAAAACAACCACTGAG 173

RESULT 6  
 US-11-136-527-4239  
 ; Sequence 4239, Application US/11136527  
 ; Publication No. US20050287570A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wyeth  
 ; APPLICANT: Mounts, William M  
 ; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
 ; FILE REFERENCE: 031896-041000 (AM101086)  
 ; CURRENT APPLICATION NUMBER: US/11/136,527  
 ; CURRENT FILING DATE: 2005-05-25  
 ; PRIOR APPLICATION NUMBER: US 60/574,294  
 ; PRIOR FILING DATE: 2005-05-26  
 ; NUMBER OF SEQ ID NOS: 362830  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 4239  
 ; LENGTH: 600  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; US-11-136-527-4239

Query Match 71.2%; Score 17.8; DB 8; Length 600;  
 Best Local Similarity 90.5%; Pred. No. 69;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



```
Query Match      69.6%; Score 17.4; DB 7; Length 21;
Best Local Similarity 94.7%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGTA 19
   |||||
Db 19 AAAAAACAACCACTGTA 1

RESULT 12
US-11-121-849-179004
; Sequence 179004, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; PRIOR FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 179004
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-179004

Query Match      68.8%; Score 17.2; DB 8; Length 25;
Best Local Similarity 86.4%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACAACCACTGAGCTGGG 25
   |||||
Db 2 AAAGCACTCCACTGAACTGGG 23

RESULT 13
US-10-750-185-34090
; Sequence 34090, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 34090
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Bovine 19866880893114
US-10-750-185-34090

Query Match      68.8%; Score 17.2; DB 7; Length 1188;
Best Local Similarity 86.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCT 22
   |||||
Db 263 AAGAACACCTTCACTGAGCT 284

RESULT 14
US-10-750-623-34090
; Sequence 34090, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 34090
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Bovine 19866880893114
US-10-750-623-34090

Query Match      68.8%; Score 17.2; DB 7; Length 1188;
Best Local Similarity 86.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCT 22
   |||||
Db 263 AAGAACACCTTCACTGAGCT 284

RESULT 15
US-10-750-185-25618
; Sequence 25618, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 25618
; LENGTH: 1324
; TYPE: DNA
; ORGANISM: Bovine 19866881355792
US-10-750-185-25618

Query Match      68.8%; Score 17.2; DB 7; Length 1324;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAACCACTGAGCT 22
   |||||
Db 1224 AAAAAACAACCTTCACTGAGCT 1245

RESULT 16
```

```
US-10-750-623-25618
; Sequence 25618, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 25618
; LENGTH: 1324
; TYPE: DNA
; ORGANISM: Bovine 19866880429771
US-10-750-623-25618

Query Match      68.8%; Score 17.2; DB 7; Length 1324;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACTGAGCT 22
   ||||| ||||| ||||| |||||
Db 1224 AAAAAAACCACTGAGCT 1245

RESULT 17
US-10-750-185-27616/c
; Sequence 27616, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27616
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Bovine 19866880429771
US-10-750-185-27616

Query Match      68.8%; Score 17.2; DB 7; Length 1826;
Best Local Similarity 86.4%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAACCACTGAGCTG 23
   ||||| ||||| ||||| |||||
Db 60 AAAAAACCACTGGGCTG 39

RESULT 18
US-10-750-623-27616/c
; Sequence 27616, Application US/10750623
; Publication No. US20050287531A1
```

---

```
US-10-750-623-27616
; Sequence 27616, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 76162
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-76162

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAACCACTGAGCTGG 25
   ||||| ||||| ||||| |||||
Db 93 AAAAAAACAACCTAGCTGG 69

RESULT 20
US-10-995-561-76164/c
; Sequence 76164, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76164
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-76164

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACCTAGCTGGG 25
||||| || |||||
Db 94 AAAAAAAACAACTTAGCTGGG 70

RESULT 21
US-10-995-561-76165/c
; Sequence 76165, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76165
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-76165

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACCTAGCTGGG 25
||||| || |||||
Db 96 AAAAAAAACAACTTAGCTGGG 72

RESULT 24
US-11-124-368A-3450/c
; Sequence 3450, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3450
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-3450

Query Match      68.0%; Score 17; DB 8; Length 201;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAACACCACCTAGCTGGG 25
||||| |||||
Db 95 AAAAAAATCCACCTAGCAGAG 71

RESULT 25
US-10-750-185-349/c
; Sequence 349, Application US/10750185
; Publication No. US200502603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```





```
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38183
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Bovine 19866880468223
US-10-750-185-38183

Query Match      68.0%; Score 17; DB 7; Length 1424;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   |||||
Db 1081 AAACAACACCACCTGAGTGAG 1105

RESULT 30
US-10-750-623-38183
; Sequence 38183, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38183
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Bovine 19866880468223
US-10-750-623-38183

Query Match      68.0%; Score 17; DB 7; Length 1424;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   |||||
Db 1081 AAACAACACCACCTGAGTGAG 1105

RESULT 31
US-11-000-463-48
; Sequence 48, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN

; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 1612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (43)..(1464)
US-11-000-463-48

Query Match      68.0%; Score 17; DB 8; Length 1612;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   |||||
Db 1570 ATAAGCACCACCTGCTGGG 1594

RESULT 32
US-11-128-061-1028
; Sequence 1028, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1028
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Cricetus griseus
US-11-128-061-1028

Query Match      68.0%; Score 17; DB 8; Length 1623;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACCTGAGCTGGG 25
   |||||
Db 267 AAAGACAAGCACCACAGAGCAGG 291
```

```
RESULT 33
US-11-128-049-1028
; Sequence 1028, Application US/111128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; TITLE OF INVENTION: MAKING AND USING SAME
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1028
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Cricetus griseus
US-11-128-049-1028

Query Match      68.0%; Score 17; DB 8; Length 1623;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25
DB 267 AAAGAGCAGCACCACTGAGCTGGG 291

RESULT 34
US-11-000-463-159
; Sequence 159, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIPACN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 159
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 666..(1499)
US-11-000-463-159

Query Match      68.0%; Score 17; DB 8; Length 1644;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25
DB 1601 ATAAAGCAGCACCACTGCCCTGGG 1625

RESULT 35
US-10-775-169-237
; Sequence 237, Application US/10775169
; Publication No. US20050287532A9
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dörner, Andrew
; APPLICANT: Trepicchio, William
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
; FILE REFERENCE: AM101080 (031896-013000)
; CURRENT APPLICATION NUMBER: US/10/775,169
; CURRENT FILING DATE: 2004-02-11
; NUMBER OF SEQ ID NOS: 5278
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 237
; LENGTH: 1673
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-169-237

Query Match      68.0%; Score 17; DB 7; Length 1673;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACACCACTGAGCTGGG 25
DB 1601 ATAAAGCAGCACCACTGCCCTGGG 1625

RESULT 36
US-10-493-909-57
; Sequence 57, Application US/10493909
; Publication No. US20060015969A1
; GENERAL INFORMATION:
; APPLICANT: LARRICK, JAMES W.
; APPLICANT: WYCOFF, KEITH L.
; TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING TOXICITY
; FILE REFERENCE: 41514-20004.01
; CURRENT APPLICATION NUMBER: US/10/493,909
; CURRENT FILING DATE: 2004-04-26
; PRIOR APPLICATION NUMBER: PCT/US01/13932
; PRIOR FILING DATE: 2001-04-28
; PRIOR APPLICATION NUMBER: 60/200,298
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-493-909-57
```

; CURRENT APPLICATION NUMBER: US/10/988,200  
 ;  
 ; CURRENT FILING DATE: 2004-11-12  
 ; PRIOR APPLICATION NUMBER: DE 103 52 900.4  
 ;

```

; PRIOR FILING DATE: 2003-11-11
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 2551
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: mab TGN1412 heavy chain
US-10-988-207-28

Query Match      68.0%; Score 17; DB 6; Length 2551;
Best Local Similarity 80.0%; Pred. No. 26+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACTGAGCTGGG 25
Db 2445 ATAAGACCCACCACTGCCCTGGG 2469

RESULT 41
US-11-000-463-630
; Sequence 630, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong B.
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 631
; LENGTH: 3229
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-000-463-631

Query Match      68.0%; Score 17; DB 8; Length 3229
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACTGAGCTGGG 25
Db 1595 ATAAGACCCACCACTGCCCTGGG 1619

RESULT 43
US-10-750-185-41816
; Sequence 41816, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41816
; LENGTH: 3451
; TYPE: DNA
; ORGANISM: Bovine 19866880792513

US-11-000-463-631
; Sequence 630, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong B.
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 630
; LENGTH: 3229
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-000-463-630

Query Match      68.0%; Score 17; DB 8; Length 3229;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCCACTGAGCTGGG 25
Db 1595 ATAAGACCCACCACTGCCCTGGG 1619

RESULT 42
US-11-000-463-631

```

US-10-750-185-41816

Query Match 68.0%; Score 17; DB 7; Length 3451;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTAGCTGGG 25

Db 2531 AAAAAACACCTAGCTGGG 2555

RESULT 44

US-10-750-623-41816  
; Sequence 41816, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41816  
; LENGTH: 3451  
; TYPE: DNA  
; ORGANISM: Bovine 19866880792513  
US-10-750-623-41816

Query Match 68.0%; Score 17; DB 7; Length 3451;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTAGCTGGG 25

Db 2531 AAAAAACACCTAGCTGGG 2555

RESULT 45

US-10-775-169-234/c  
; Sequence 234, Application US/10775169  
; Publication No. US20050287532A9  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dornet, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 234  
; LENGTH: 35962  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-234

Query Match 68.0%; Score 17; DB 7; Length 35962;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTAGCTGGG 25

Db 9691 AAAAAACCAAAATAGCTGGG 9667

RESULT 46

US-10-995-561-13472/c  
; Sequence 13472, Application US/10995561  
; Publication No. US2005027054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13472  
; LENGTH: 50529  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(50529)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-995-561-13472

Query Match 68.0%; Score 17; DB 7; Length 50529;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTAGCTGGG 25

Db 3800 AAAAAACACCACTAGCTGGG 3776

RESULT 47

US-10-775-169-193  
; Sequence 193, Application US/10775169  
; Publication No. US20050287532A9  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dornet, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 193  
; LENGTH: 110608  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-193

Query Match 68.0%; Score 17; DB 7; Length 110608;  
Best Local Similarity 80.0%; Pred. No. 3.9e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACCACTAGCTGGG 25

Db 95809 AAAAAACACCAAAATAGCTGGG 95833

RESULT 48

US-11-112-908-24  
; Sequence 24, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:

; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 24  
; LENGTH: 150314  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-24

Query Match 68.0%; Score 17; DB 8; Length 150314;  
Best Local Similarity 80.0%; Pred. No. 4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACACCACTGAGCTGGG 25  
|||||  
DB 97203 AAAAAACAAAAAATAATTAGCTGGG 97227

RESULT 49  
US-11-172-274-1/c  
; Sequence 1, Application US/11/172274  
; Publication No. US20060014253A1  
; GENERAL INFORMATION:  
; APPLICANT: University Of New Hampshire  
; APPLICANT: Sower, Stacia A  
; APPLICANT: Silver, Matt  
; TITLE OF INVENTION: Novel Polynucleotides Encoding Lamprey GnRH-III  
; FILE REFERENCE: 9815/59339  
; CURRENT APPLICATION NUMBER: US/11/172,274  
; CURRENT FILING DATE: 2005-06-30  
; PRIOR APPLICATION NUMBER: US/10/170,096  
; PRIOR FILING DATE: 2002-06-12  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 790  
; TYPE: DNA  
; ORGANISM: g. australis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: (199)..()  
; OTHER INFORMATION:  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (124)..(405)  
; OTHER INFORMATION:  
US-11-172-274-1

Query Match 67.2%; Score 16.8; DB 8; Length 790;  
Best Local Similarity 90.0%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAACACCACTGAG 20  
|||||  
DB 59 AAAAAATCAACCAACTGAG 40

RESULT 50  
US-10-775-169-345/c  
; Sequence 345, Application US/10775169

; Publication No. US20050287532A9  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dorner, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 345  
; LENGTH: 128978  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-345  
Query Match 67.2%; Score 16.8; DB 7; Length 128978;  
Best Local Similarity 90.0%; Pred. No. 4.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAACACCACTGAG 20  
|||||  
DB 78466 AAAAAACACCACTGAG 78447  
Search completed: February 3, 2006, 16:19:47  
Job time : 346.111 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-5  
Perfect score: 25  
Sequence: 1 aaaaaaatgaccactggctgt 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_ets.\*  
11: gb\_ey.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23.4	93.6	187589	9	AC153565	AC153565 Mus muscu
C 2	23.4	93.6	228194	9	AC153893	AC153893 Mus muscu
C 3	20.8	83.2	57248	8	HS15E1	AL021946 Human DNA
C 4	20.8	83.2	129794	8	AL353812	AL353812 Human DNA
C 5	20.8	83.2	136515	8	AL499609	AL499609 Human DNA
C 6	20.8	83.2	153605	14	AC019360	AC019360 Homo sapi
C 7	20.8	83.2	168354	14	AC133867	AC133867 Mus muscu
C 8	20.8	83.2	170429	9	AC142146	AC142146 Mus muscu
C 9	20.8	83.2	171970	14	AL590454	AL590454 Homo sapi
C 10	20.8	83.2	177654	14	AC025988	AC025988 Homo sapi
C 11	20.8	83.2	183861	14	HS75N14	297199 Homo sapien
C 12	20.8	83.2	187546	14	AC129236	AC129236 Homo sapi
C 13	20.2	80.8	601	10	BV293157	BV293157 S232P6532
C 14	20.2	80.8	40745	14	AC165059	AC165059 Phakopsor
C 15	20.2	80.8	57175	14	AC100243	AC100243 Mus muscu
C 16	20.2	80.8	62356	14	AC111185	AC111185 Homo sapi
C 17	20.2	80.8	87871	8	AL356141	AL356141 Human DNA
C 18	20.2	80.8	134231	14	AC154170	AC154170 Alligator

AC090735	Homo sapi	8	AC090735	80.8	149089	20.2	19
AC012339	Homo sapi	8	AC012339	80.8	179818	20.2	C 20
AC068709	Homo sapi	14	AC068709	80.8	180583	20.2	C 21
AC013562	Homo sapi	8	AC013562	80.8	188755	20.2	C 22
AC023376	Homo sapi	14	AC023376	80.8	189963	20.2	C 23
AC074145	Mus muscu	14	AC074145	80.8	208161	20.2	C 24
AL683818	Mouse DNA	9	AL683818	80.8	210718	20.2	C 25
AC094552	Rattus no	14	AC094552	80.8	246315	20.2	C 26
AC094299	Rattus no	14	AC094299	80.8	249661	20.2	C 27
AC146241	Pan trogl	8	AC146241	79.2	160620	19.8	C 28
AC166995	Mus muscu	14	AC166995	79.2	170524	19.8	C 29
AC073470	Homo sapi	14	AC073470	79.2	171673	19.8	C 30
AC018728	Homo sapi	8	AC018728	79.2	181935	19.8	C 31
AC140950	Pan trogl	8	AC140950	79.2	190084	19.8	C 32
CT010460	Mus muscu	14	CT010460	79.2	232469	19.8	C 33
AK072591	Oryza sat	15	AK072591	77.6	2346	19.4	C 34
Continuation (262		15	AP008212_261	77.6	110000	19.4	C 35
AP003517	Oryza sat	15	AP003517	77.6	151636	19.4	C 36
CO337709	Sequence	6	CO337709	77.6	175047	19.4	C 37
AP005453	Oryza sat	15	AP005453	76.8	600	19.2	C 38
AC117517	Homo sapi	15	AC117517	76.8	101607	19.2	C 39
AP004334	Oryza sat	15	AP004334	76.8	1310	19.2	C 40
AC108692	Homo sapi	14	AC108692_0	76.8	110000	19.2	C 41
Continuation (263		15	AP008213_262	76.8	110000	19.2	C 42
AC152963	Ikafugu	14	AC152963	76.8	111804	19.2	C 43
AC117517	Homo sapi	8	AC117517	76.8	121628	19.2	C 44
AC144797	Mus muscu	9	AC144797	76.8	129045	19.2	C 45
AC135550	Pan trogl	14	AC135550	76.8	129968	19.2	C 46
AC144765	Medicago	15	AC144765	76.8	130836	19.2	C 47
AC145828	Pan trogl	8	AC145828	76.8	139712	19.2	C 48
AC092338	Homo sapi	8	AC092338	76.8	143126	19.2	C 49
AC027584	Homo sapi	8	AC027584	76.8	144003	19.2	C 50
AC156475	Bos tauru	14	AC156475	76.8	144371	19.2	C 51
AC149788	Pan trogl	8	AC149788	76.8	171501	19.2	C 52
AC069059	Homo sapi	14	AC069059	76.8	174631	19.2	C 53
AC107366	Mus muscu	14	AC107366	76.8	174714	19.2	C 54
AC116892	Mus muscu	9	AC116892	76.8	176340	19.2	C 55
AC009178	Homo sapi	8	AC009178	76.8	178097	19.2	C 56
AC116108	Mus muscu	9	AC116108	76.8	179033	19.2	C 57
AC140183	Mus muscu	9	AC140183	76.8	181029	19.2	C 58
AC104328	Mus muscu	9	AC104328	76.8	183187	19.2	C 59
AC133553	Homo sapi	14	AC133553	76.8	185887	19.2	C 60
BX537123	Zebrafish	5	BX537123	76.8	202654	19.2	C 61
AC160633	Mus muscu	9	AC160633	76.8	203628	19.2	C 62
AC118214	Mus muscu	14	AC118214	76.8	204736	19.2	C 63
AC128722	Rattus no	14	AC128722	76.8	210126	19.2	C 64
AC113200	Mus muscu	14	AC113200	76.8	215792	19.2	C 65
AC15154	Mus muscu	14	AC15154	76.8	223524	19.2	C 66
AC109972	Rattus no	14	AC109972	76.8	224397	19.2	C 67
AC162461	Mus muscu	5	AC162461	76.8	226685	19.2	C 68
AC094429	Rattus no	14	AC094429	76.8	231424	19.2	C 69
AC130152	Rattus no	14	AC130152	76.8	235331	19.2	C 70
AC105697	Rattus no	14	AC105697	76.8	244170	19.2	C 71
AC103346	Rattus no	14	AC103346	76.8	244853	19.2	C 72
AC109397	Rattus no	14	AC109397	76.8	246106	19.2	C 73
AC158061	Bos tauru	14	AC158061	76.8	256373	19.2	C 74
AC109412	Rattus no	14	AC109412	76.8	265598	19.2	C 75
AC164651	Ornithorh	14	AC164651	76.8	319131	19.2	C 76
CQ581591	Sequence	6	CQ581591	76.0	182043	19	C 77
AC009984	Drosophil	14	AC009984	75.2	6596	18.8	C 78
AC013977	Drosophil	14	AC013977	75.2	102468	18.8	C 79
AC162728	Loxodonta	14	AC162728	75.2	106584	18.8	C 80
AC162734	Loxodonta	14	AC162734	75.2	135532	18.8	C 81
AC008140	Drosophil	2	AC008140	75.2	138174	18.8	C 82
AC152466	Daaypus n	14	AC152466	75.2	140973	18.8	C 83
AC162475	Rattus no	14	AC162475	75.2	144433	18.8	C 84
AC152479	Daaypus n	14	AC152479	75.2	145370	18.8	C 85
AC009219	Drosophil	2	AC009219	75.2	152607	18.8	C 86
AC009314	Homo sapi	8	AC009314	75.2	159007	18.8	C 87
AC161706	Otolemur	14	AC161706	75.2	161788	18.8	C 88
AC087651	Homo sapi	14	AC087651	75.2	164715	18.8	C 89
AC024725	Homo sapi	14	AC024725	75.2	171748	18.8	C 90
		14		75.2	197352	18.8	C 91

[illegible]

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

```

source
1. .187589
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="10"
   /clones="RP23-228H2"

misc_feature
1. .44609
   /notes="overlaps bases 183586. .228194 of clone AC153893"
   /function="Clone overlap"

repeat_region
135. .225
   /rpt_family="PB1D7"

repeat_region
317. .555
   /rpt_family="B4A"

repeat_region
614. .656
   /rpt_family="(CA)n"

repeat_region
1502. .1538
   /rpt_family="(CAG)n"

repeat_region
1894. .2386
   /rpt_family="L1MB4"

repeat_region
2850. .2884
   /rpt_family="(TAAA)n"

repeat_region
3563. .4822
   /rpt_family="ORR1D-int"

repeat_region
4823. .4866
   /rpt_family="(TG)n"

repeat_region
4867. .4892
   /rpt_family="ORR1D-int"

repeat_region
4893. .5214
   /rpt_family="ORR1D"

repeat_region
5768. .5823
   /rpt_family="(GGAA)n"

repeat_region
5824. .5983
   /rpt_family="(GGGAA)n"

repeat_region
6289. .6509
   /rpt_family="B3"

repeat_region
7086. .7127
   /rpt_family="(CA)n"

repeat_region
8477. .8497
   /rpt_family="AT-rich"

repeat_region
complement(8550. .8740)
   /rpt_family="B3A"

repeat_region
9885. .9929
   /rpt_family="(GGGA)n"

repeat_region
complement(11090. .11407)
   /rpt_family="ORR1B1"

repeat_region
complement(11553. .11706)
   /rpt_family="RSINE1"

repeat_region
complement(11717. .11823)
   /rpt_family="PB1D10"

repeat_region
complement(11744. .11843)
   /rpt_family="MT2B"

repeat_region
14853. .14997
   /rpt_family="B4A"

repeat_region
16135. .16413
   /rpt_family="B4"

repeat_region
16443. .16496
   /rpt_family="(CA)n"

repeat_region
17042. .17267
   /rpt_family="MTD"

repeat_region
18539. .18620
   /rpt_family="AT-rich"

repeat_region
21422. .21471
   /rpt_family="(TG)n"

repeat_region
21626. .21658
   /rpt_family="(TCC)n"

repeat_region
complement(21684. .21852)

```

```

repeat_region
22693. .22785
   /rpt_family="B1F"

repeat_region
22976. .23025
   /rpt_family="B4A"

repeat_region
23121. .23171
   /rpt_family="(CATA)n"

repeat_region
24199. .24296
   /rpt_family="PB1D10"

repeat_region
24348. .24369
   /rpt_family="AT-rich"

repeat_region
25089. .25169
   /rpt_family="B1_Mur2"

repeat_region
25170. .25227
   /rpt_family="(GAAA)n"

repeat_region
25230. .25388
   /rpt_family="(GGAA)n"

repeat_region
26120. .26214
   /rpt_family="B1_Mur1"

repeat_region
26335. .26517
   /rpt_family="B3"

repeat_region
26574. .26761
   /rpt_family="L1MB8"

repeat_region
complement(26775. .26989)
   /rpt_family="B3"

repeat_region
27249. .27389
   /rpt_family="(CATATA)n"

repeat_region
27801. .28024
   /rpt_family="B4A"

repeat_region
28461. .28517
   /rpt_family="(CA)n"

repeat_region
complement(29233. .29348)
   /rpt_family="MIR"

repeat_region
29701. .29831
   /rpt_family="B1F"

repeat_region
30002. .30040
   /rpt_family="A-rich"

repeat_region
30090. .30240
   /rpt_family="B1_Mur3"

repeat_region
30581. .30632
   /rpt_family="(CA)n"

repeat_region
31143. .31165
   /rpt_family="AT-rich"

repeat_region
31166. .31276
   /rpt_family="PB1D10"

repeat_region
31279. .31438
   /rpt_family="B3A"

```

Query Match 93.6%; Score 23.4; DB 9; Length 187589;  
 Best Local Similarity 96.0%; Pred. No. 3.4;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCTGT 25  
 |||||  
 Db 32277 AAAAAAATGACCAACTGGGCTGT 32253

## RESULT 2

AC153893/c  
 LOCUS AC153893 228194 bp DNA linear ROD 29-JAN-2005  
 DEFINITION Mus musculus 10 BAC RP23-124O20 (Roswell Park Cancer Institute  
 (C57BL/6J Female) Mouse BAC Library) complete sequence.  
 ACCESSION AC153893  
 VERSION AC153893.3 GI:58257766  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 228194)  
 AUTHORS Muzny,D., Adams,C., Agbai II,O., Allen,C., Alabrooks,S., Archer,P.,  
 Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,

Beraducci, A., Biswal, K., Blyth, P., Bonham, H., Buhay, C., Burch, P., Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I., Chacko, J., Chahrour, M., Chavez, D., Chen, A., Chen, G., Chen, R., Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A., Curry, S., Dai, W., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, C., Dziulda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P., Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, N., Garner, T., Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haerlen, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hogue, M., Hollins, B., Howell, L. T., Hui, K., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., LeGall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensheva, L., Lozada, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadao, C., Metzker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazarith, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Kott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Raigh, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Usmani, K., Vargo, C., Verdusco, D., Villaseana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

TITLE  
Direct Submission

JOURNAL  
REFERENCE  
2 (bases 1 to 228194)

AUTHORS  
Worley, K.C.

TITLE  
JOURNAL  
Submitted (18-DEC-2004) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 228194)

REFERENCE  
Worley, K.C.TITLE  
Direct Submission

JOURNAL  
Submitted (08-JAN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 228194)

REFERENCE  
Worley, K.C.TITLE  
Direct Submission

JOURNAL  
Submitted (28-JAN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 228194)

REFERENCE  
Worley, K.C.TITLE  
Direct Submission

JOURNAL  
Submitted (29-JAN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 28, 2005 this sequence version replaced gi:57334844.

Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

## source

location/Qualifiers  
1..228194  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-124020"  
1..1267  
/rpt\_family="L1Md\_T"  
complement(5200..5414)  
/rpt\_family="URR1A"  
5482..5658  
/rpt\_family="TTC)n"  
5573  
/note="polymorphism: T/C, MSZWP/outside reads"  
/function="polymorphic site"  
5659..5762  
/rpt\_family="TCC)n"  
5771..5868  
/rpt\_family="B3"  
6322..6357  
/rpt\_family="GA)n"  
6857..6912  
/rpt\_family="GAAA)n"  
8508..8530  
/rpt\_family="AT-rich"  
10134..10232  
/rpt\_family="CA)n"  
10776..11064  
/rpt\_family="B4A"  
11081..11138  
/rpt\_family="CA)n"  
complement(12651..12874)  
/rpt\_family="URR1A"  
13728..13789  
/rpt\_family="GGGA)n"  
14331..14353  
/rpt\_family="TGGG)n"  
complement(15471..15529)  
/rpt\_family="RMR30"  
16544..16678  
/rpt\_family="GGGAGA)n"  
16738..16791  
/rpt\_family="CA)n"  
17125..17440  
/rpt\_family="ORR1B1"  
17477..17607  
/rpt\_family="RSINE1"  
17613..17654  
/rpt\_family="CA)n"  
18510..18656  
/rpt\_family="RSINE1"  
19893..19998  
/rpt\_family="PB1D10"  
20005..20063  
/rpt\_family="A-rich"  
20782..20976  
/rpt\_family="URR1A"  
21547..21827  
/rpt\_family="B4A"  
complement(23063..23205)  
/rpt\_family="B1\_MuB1"  
25349..25375  
/rpt\_family="TA)n"  
complement(26639..26848)  
/rpt\_family="B3"  
28245..28498  
/rpt\_family="B4A"  
complement(29029..29107)  
/rpt\_family="B4A"

repeat\_region

repeat\_region

repeat\_region

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```

repeat_region 29544..29579
/rpt_family="AT_rich"
repeat_region 29582..29657
/rpt_family="ID4_"
/rpt_family="ID4_"
repeat_region 29790..29822
/rpt_family="(GA)n"
repeat_region 29828..29927
/rpt_family="(GGA)n"
repeat_region 30730..30848
/rpt_family="PB1D9"
repeat_region 30889..30980
/rpt_family="B1_Mus1"
repeat_region 30959..31002
/rpt_family="ID_B1"
repeat_region 31018..31117
/rpt_family="(GGGGA)n"
repeat_region complement(31293..31412)
/rpt_family="B1_Mur3"
repeat_region complement(32054..32268)
/rpt_family="L1MB7"
repeat_region complement(33129..33376)
/rpt_family="Lx6"
repeat_region 34113..34143
/rpt_family="AT_rich"
repeat_region complement(34145..34346)
/rpt_family="B3"
repeat_region 35847..35879
/rpt_family="AT_rich"
repeat_region 37543..37641
/rpt_family="MLT1B"
repeat_region 38638..38770
/rpt_family="L1MA9"
repeat_region 38805..38843
/rpt_family="(CA)n"
repeat_region 38844..38891
/rpt_family="(GA)n"
repeat_region 38916..39185
/rpt_family="L1MA9"
repeat_region 39194..39305
/rpt_family="PB1"

```

Query Match 93.6%; Score 23.4; DB 9; Length 228194;  
 Best Local Similarity 96.0%; Pred. No. 3.4;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 AAAAAAATGACCACTGGGCTGT 25
|||||
Db 215862 AAAAAAATGACCACTGGGCTGT 215838
|||||

```

```

RESULT 3
HS15E1 57248 bp DNA linear PRI 16-APR-2005
LOCUS Human DNA sequence from clone XX-15E1 on chromosome 12, complete
DEFINITION
ACCESSION AL021546
VERSION AL021546.1 GI:2826890
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 57248)
AUTHORS Murphy, L.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

```

Contact: humquery@sanger.ac.uk

-----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 En:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/IMPORTANT: This  
 sequence is not the entire insert of clone XX-15E1 It may be  
 shorter because we sequence overlapping sections only once, except  
 for a short overlap.

The true left end of clone RP1-75N14 is at 55991 in this sequence.  
 The true right end of clone XX-166H1 is at 1098 in this sequence.

#### FEATURES

```

Source
Location/Qualifiers
1..57248
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="XX-15E1"
1..97
/note="L1MC1 repeat: matches 6233..6324 of consensus"
repeat_region 132..438
/note="AluYa8 repeat: matches 1..302 of consensus"
repeat_region 438..718
/note="AluJo repeat: matches 1..281 of consensus"
repeat_region 745..1035
/note="AluX repeat: matches 1..292 of consensus"
repeat_region 1039..1335
/note="AluX repeat: matches 3..299 of consensus"
repeat_region 1447..1552
/note="MIR repeat: matches 100..212 of consensus"
repeat_region 1830..2117
/note="AluSq repeat: matches 1..291 of consensus"
repeat_region 2118..2420
/note="AluY repeat: matches 8..307 of consensus"
repeat_region 2748..2976
/note="MIR repeat: matches 28..262 of consensus"
repeat_region 3027..3130
/note="52 copies 2 mer cc 60% conserved"
repeat_region 3215..3507
/note="AluY repeat: matches 1..307 of consensus"
repeat_region 3551..3862
/note="AluX repeat: matches 1..312 of consensus"
repeat_region 3921..4018
/note="L2 repeat: matches 2597..2706 of consensus"
repeat_region 4036..4342
/note="AluJo repeat: matches 1..292 of consensus"
misc_feature complement(4343..4423)
/note="match: GSS: Em:AQ621351"
repeat_region 4424..4487
/note="MER5A repeat: matches 45..116 of consensus"
repeat_region 4573..4874
/note="AluX repeat: matches 1..306 of consensus"
repeat_region 4948..5076
/note="FLAM C repeat: matches 5..133 of consensus"
repeat_region 5082..5150
/note="L2 repeat: matches 2404..2471 of consensus"
repeat_region 5713..5901
/note="MIR repeat: matches 13..194 of consensus"

```

```
repeat_region 5910. .6217
/note="LJMC5 repeat: matches 7444. .7758 of consensus"
repeat_region 6218. .6245
/note="14 copies 2 mer gt 89% conserved"
repeat_region 6247. .6411
/note="AluJo repeat: matches 109. .285 of consensus"
repeat_region 6412. .6716
/note="AluY repeat: matches 2. .306 of consensus"
repeat_region 6722. .7039
/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region 7040. .7165
/note="AluJo repeat: matches 11. .105 of consensus"
repeat_region 7183. .7335
/note="LJMA4A repeat: matches 5998. .6277 of consensus"
repeat_region 7371. .7592
/note="AluYb repeat: matches 1. .228 of consensus"
repeat_region 7669. .7759
/note="L2 repeat: matches 2629. .2714 of consensus"
repeat_region 7760. .8058
/note="AluY repeat: matches 2. .297 of consensus"
repeat_region 8059. .8087
/note="Alu repeat: matches 117. .145 of consensus"
repeat_region 8088. .8161
/note="AluS repeat: matches 231. .304 of consensus"
repeat_region 8179. .8481
/note="AluSp repeat: matches 1. .304 of consensus"
repeat_region 8489. .8771
/note="AluJo repeat: matches 1. .288 of consensus"
repeat_region 8990. .9055
/note="33 copies 2 mer ta 72% conserved"
repeat_region 9057. .9219
/note="AluSg/x repeat: matches 134. .296 of consensus"
repeat_region 9221. .9308
/note="L2 repeat: matches 2637. .2725 of consensus"
repeat_region 9339. .9596
/note="L2 repeat: matches 2135. .2419 of consensus"
repeat_region 9601. .9735
/note="AluYb repeat: matches 1. .135 of consensus"
repeat_region 9736. .9926
/note="AluSx repeat: matches 1. .210 of consensus"
repeat_region 9927. .10144
/note="AluSx repeat: matches 1. .216 of consensus"
repeat_region 10145. .10252
/note="AluSx repeat: matches 210. .312 of consensus"
repeat_region 10483. .10556
/note="L2 repeat: matches 2624. .2695 of consensus"
repeat_region 10601. .10894
/note="AluY repeat: matches 1. .297 of consensus"
repeat_region 10895. .11031
/note="AluYb repeat: matches 1. .137 of consensus"
repeat_region 11068. .11200
/note="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 11525. .11682
/note="MIR repeat: matches 92. .262 of consensus"
repeat_region 11733. .12027
/note="AluSg repeat: matches 1. .295 of consensus"
repeat_region 12027. .12306
/note="AluYb repeat: matches 1. .281 of consensus"
repeat_region 12426. .12723
/note="AluSx repeat: matches 1. .298 of consensus"
repeat_region 13132. .13335
/note="MIR repeat: matches 6. .212 of consensus"
repeat_region 13445. .13594
/note="AluSg/x repeat: matches 155. .297 of consensus"
repeat_region 13595. .13895
/note="AluSg repeat: matches 1. .292 of consensus"
repeat_region 13896. .14197
/note="AluSx repeat: matches 1. .305 of consensus"
repeat_region 14198. .14222
/note="AluSg/x repeat: matches 113. .155 of consensus"
repeat_region 14488. .14785
/note="AluSg repeat: matches 3. .300 of consensus"
repeat_region 14804. .14854

/note="MIR repeat: matches 210. .260 of consensus"
14859. .14925
/note="MIR repeat: matches 2. .69 of consensus"
14926. .15105
/note="AluJo repeat: matches 1. .183 of consensus"
15111. .15398
/note="AluSx repeat: matches 6. .292 of consensus"
15466. .15525
/note="MIR repeat: matches 14. .85 of consensus"
15526. .15642
/note="FLAM A repeat: matches 20. .136 of consensus"
15858. .15949
/note="MIR repeat: matches 23. .122 of consensus"
15967. .16270
/note="AluSx repeat: matches 1. .305 of consensus"
16271. .16571
/note="AluYb repeat: matches 1. .300 of consensus"
16720. .16843
/note="FLAM C repeat: matches 7. .134 of consensus"
16877. .16902
/note="13 copies 2 mer aa 100% conserved"
16903. .17218
/note="AluSg repeat: matches 1. .313 of consensus"
17307. .17556
/note="L2 repeat: matches 2485. .2750 of consensus"
17895. .20523
/gene="COX6A2"
Join(17895. .18023,18173. .18315,20248. .20523)
/gene="COX6A2"
/product="15E1.4 (Cytochrome C Oxidase Polypeptide
Vfa-liver precursor (EC 1.9.3.1))"
/evidence=not experimental
Join(17921. .18023,18173. .18315,20248. .20331)

CDS

Query Match 83.2%; Score 20.8; DB 8; Length 57248;
Best Local Similarity 91.7%; Pred. No. 63;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCACTGGGCTG 24
Db 42908 AAAAAACATGTCACACTGGTCTG 42931

RESULT 4
AL353812/c
LOCUS AL353812 129794 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-384D7 on chromosome 20 Contains
the 5' end of the C20orf22 for chromosome 20 open reading frame 22
(DKFZp434P106), the PPIAP2 gene for peptidylprolyl isomerase A
(cyclophilin A) pseudogene 2, the 5' end of the gene for a novel
protein (K1AA0186) and three CpG islands, complete sequence.
ACCESSION AL353812
VERSION AL353812.13 GI:8670913
KEYWORDS HTG; C20orf22; CpG island; cyclophilin; DKFZ434P106; K1AA0186;
peptidylprolyl isomerase; PPIAP2.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 129794)
AUTHORS Dunn, M.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jun 22, 2000 this sequence version replaced gi:8670605.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
```



chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>  
 RP11-384D7 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

Location/Qualifiers

1..129794

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="RZPD:RPCIB753D07384"

/db\_xref="taxon:9606"

/chromosome="20"

/clone="RP11-384D7"

/clone\_lib="RPC1-11.2"

100

/note="Clone right end: RP5-965G21"

join(complement(84492..84955), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894), complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:69576..70262))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-001"  
 join(complement(84492..84955), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894), complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:64121..64408))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-002"  
 join(complement(84492..84955), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894), complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:69576..70262))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-001"  
 /note="match: ESTs: AY722769 AL527984 AL527985 AL530814 AL53225 AL568353 AL577050 AW414190 AW949122 BE019167 BE612777 BE737507 BE757201 BE876190 BE905978 BE908609 BG031660 BG297499 BG386466 BG800424 BI160243 BI600796 BI755474 BM544474 BM663934 BM763102 BM793216 BM976653 BQ029143 BQ368701 BQ574269  
 match: cDNAs: AL117442 BC002263"  
 join(complement(84492..84955), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894),

mRNA

complement(1945..2024), complement(812..894), complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /notes="match: ESTs: BI765350"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /locus\_tag="RP5-965G21.2-007"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554),

complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:64121..64408))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-001"  
 /note="match: ESTs: AA807832 AA946928 AI015541 AI244231 AI384054 AI401371 AI623375 AU534127 AU545563 AU549597 AU565047 AW303865 BE221580 BF346499 BG386466 BG719480 BI160243 BI822277 BI918919 BM668345 BM692995 N34825 N52041  
 match: cDNAs: BC014049"

join(complement(84492..84682), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894), complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:64351..64408))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-001"

/standard\_name="OTTHUMP0000030493"

/note="match: proteins: Q96CR1"

/codon\_start=1

/protein\_id="CAI13762.1"

/db\_xref="GI:55958487"

/translation="MRKRTEPVALEHERCAAAGSSSSGSAALDADCRKLKQNLRLTG  
 PAAAPRCAADAGMKRALGRKGVMLRLKILFCVLGLYIAIPFLIKLCPGIAKLIF  
 LNFVRVYFDILKPKDQGLNHTCNYYLOPEDVTIGVHVTVPVAVWKNQAGKQDMWY  
 EDALASSHPILLYLHGNACTRGDHRVELYKVLSSIGYHVVTDFYRWGDSVGTPSER  
 GMTYDALHVFDMIKARSGNDPNYINGHSLGTGVTATNLVRLCERETPPDALLIESPPT  
 NTRREAKSHFVSIVYIFPGDFWFLDPTSSGKGFANDENVKHISCPLLILHAEDDP  
 VYFQGRKLYSIAAPARSFRDFKQVFPFHSDLYGRHKYIKSPSLRILRFPQGGG  
 SSPDPSMWSELY"  
 join(complement(84492..84682), complement(33206..33330), complement(17304..17409), complement(14178..14297), complement(11027..11057), complement(8904..8949), complement(3425..3554), complement(2436..2473), complement(1945..2024), complement(812..894), complement(AL121772.19:72928..73006), complement(AL121772.19:71597..71724), complement(AL121772.19:70223..70262))  
 /gene="C20orf22"

CDS

/locus\_tag="RP5-965G21.2-001"  
 /standard\_name="OTTHUMP0000030492"  
 /note="match: proteins: Q99M06"  
 /codon\_start=1  
 /protein\_id="CAI13763.1"  
 /db\_xref="GI:55958488"

/translation="MRKRTEPVALEHERCAAAGSSSSGSAALDADCRKLKQNLRLTG  
 LNFVRVYFDILKPKDQGLNHTCNYYLOPEDVTIGVHVTVPVAVWKNQAGKQDMWY  
 EDALASSHPILLYLHGNACTRGDHRVELYKVLSSIGYHVVTDFYRWGDSVGTPSER  
 GMTYDALHVFDMIKARSGNDPNYINGHSLGTGVTATNLVRLCERETPPDALLIESPPT  
 NTRREAKSHFVSIVYIFPGDFWFLDPTSSGKGFANDENVKHISCPLLILHAEDDP  
 VYFQGRKLYSIAAPARSFRDFKQVFPFHSDLYGRHKYIKSPSLRILRFLREFGKS  
 EPEHQH"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554), 8904..8949,11027..11057,14178..14297,17304..17409,33206..33330,33674..33710))  
 /gene="C20orf22"

/locus\_tag="RP5-965G21.2-006"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554), 8904..8949,11027..11057,14178..14297,17304..17409,33206..33330,33674..33710))  
 /gene="C20orf22"

gene

/locus\_tag="RP5-965G21.2-006"  
 /note="match: ESTs: BI765350"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /locus\_tag="RP5-965G21.2-007"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554),

mRNA

complement(1945..2024), complement(812..894), complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /notes="match: ESTs: BI765350"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /locus\_tag="RP5-965G21.2-007"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554),

gene

complement(1945..2024), complement(812..894), complement(join(812..894,1945..2024,2436..2473,3425..3554), 4179..4451))  
 /locus\_tag="RP5-965G21.2-007"  
 complement(join(812..894,1945..2024,2436..2473,3425..3554),

mRNA





Boguelavkiy, I., Boukhgalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
Galgan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
McElrim, J., Meneus, L., Mhova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neil, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# TITLE

## JOURNAL

## COMMENT

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 8, 2000 this sequence version replaced gi:6715815.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5690

Center clone name: 777\_M\_5

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 138398 bases at least Q40

Consensus quality: 145295 bases at least Q30

Consensus quality: 147783 bases at least Q20

Insert size: 171000; agarose-fp

Insert size: 150305; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1357: contig of 1357 bp in length  
\* 1358 1457: gap of 100 bp  
\* 1359 2566: contig of 1109 bp in length  
\* 2567 2666: gap of 100 bp  
\* 2667 3828: contig of 1162 bp in length  
\* 3829 3929: gap of 100 bp  
\* 3929 5158: contig of 1229 bp in length  
\* 5158 5257: gap of 100 bp  
\* 5257 6995: contig of 1738 bp in length  
\* 6995 7095: gap of 100 bp  
\* 7095 8163: contig of 1068 bp in length  
\* 8163 8263: gap of 100 bp  
\* 8263 9738: contig of 1475 bp in length  
\* 9738 9839: gap of 100 bp  
\* 9839 12015: contig of 2177 bp in length  
\* 12015 12115: gap of 100 bp  
\* 12115 14153: contig of 2038 bp in length  
\* 14153 14253: gap of 100 bp  
\* 14253 14254 15947: contig of 1694 bp in length

\* 15948 16047: gap of 100 bp  
\* 16048 17790: contig of 1743 bp in length  
\* 17791 17890: gap of 100 bp  
\* 17891 20931: contig of 3041 bp in length  
\* 20932 21031: gap of 100 bp  
\* 21032 24518: contig of 3487 bp in length  
\* 24519 24619: gap of 100 bp  
\* 24619 27964: contig of 3346 bp in length  
\* 27965 28064: gap of 100 bp  
\* 28065 30904: contig of 2740 bp in length  
\* 30905 34188: contig of 3284 bp in length  
\* 34189 34288: gap of 100 bp  
\* 34289 38917: contig of 4629 bp in length  
\* 38918 43237: gap of 100 bp  
\* 39018 43237: contig of 4220 bp in length  
\* 43238 47566: contig of 4229 bp in length  
\* 43338 47567 47666: gap of 100 bp  
\* 47667 51301: contig of 3635 bp in length  
\* 51302 51401: gap of 100 bp  
\* 51402 55847: contig of 4446 bp in length  
\* 55848 55947: gap of 100 bp  
\* 55948 60782: contig of 4835 bp in length  
\* 60783 60882: gap of 100 bp  
\* 60883 65110: contig of 4128 bp in length  
\* 65111 67924: contig of 2814 bp in length  
\* 67925 68024: gap of 100 bp  
\* 68025 72012: contig of 3988 bp in length  
\* 72013 72112: gap of 100 bp  
\* 72113 78507: contig of 6395 bp in length  
\* 78508 78607: gap of 100 bp  
\* 78608 85236: contig of 6629 bp in length  
\* 85237 85336: gap of 100 bp  
\* 85337 91732: contig of 6396 bp in length  
\* 91733 91832: gap of 100 bp  
\* 91833 98668: contig of 6836 bp in length  
\* 98669 98768: gap of 100 bp  
\* 98769 107846: contig of 9078 bp in length  
\* 107847 107946: gap of 100 bp  
\* 107947 116112: contig of 8166 bp in length  
\* 116113 116212: gap of 100 bp  
\* 116213 124989: contig of 8777 bp in length  
\* 124990 125089: gap of 100 bp  
\* 125090 135741: contig of 10652 bp in length  
\* 135742 135841: gap of 100 bp  
\* 135842 153605: contig of 17764 bp in length.

FEATURES  
Location/Qualifiers  
source  
1. .153605  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="12"  
/clone="RP11-777M5"  
/clone\_lib="RP11-777M5 Human Male BAC"  
misc\_feature  
1. .1357  
/note="assembly\_fragment"  
1358. .1457  
/estimated\_length=100  
misc\_feature  
1458. .2566  
/note="assembly\_fragment"  
2567. .2666  
/estimated\_length=100  
2667. .3828  
/note="assembly\_fragment"  
3829. .3928  
/estimated\_length=100  
misc\_feature  
3929. .5157  
/note="assembly\_fragment"  
5158. .5257  
/estimated\_length=100  
gap

```

misc_feature      5258..6995
                  /note="assembly_fragment"
gap              6996..7095
                  /estimated_length=100
misc_feature      7096..8163
                  /note="assembly_fragment"
gap              8164..8263
                  /estimated_length=100
misc_feature      8264..9738
                  /note="assembly_fragment"
gap              9739..9838
                  /estimated_length=100
misc_feature      9839..12015
                  /note="assembly_fragment"
gap              12016..12115

Query Match      83.2%; Score 20.8; DB 14; Length 153605;
Best Local Similarity 91.7%; Pred. No. 61;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCACTGGGCTG 24
    ||||| ||||| ||||| ||||| |||||
Db 103768 AAAAAACAATGTCCAACTGGTCTG 103745

RESULT 7
AC133867 169354 bp DNA linear HTG 04-MAR-2003
LOCUS Mus musculus clone RP24-211L7, WORKING DRAFT SEQUENCE, 11 unordered
DEFINITION
PIECES.
AC133867
VERSION AC133867.3 GI:28827899
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM
EuKaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 169354)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-211L7
Unpublished
2 (bases 1 to 169354)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 169354)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,P., Corum,B., DeArellano,K., Erickson,J.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

```

```

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 4, 2003 this sequence version replaced gi:28416243.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21215
Center clone name: 211.L.7
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165421 bases at least Q40
Consensus quality: 166530 bases at least Q30
Consensus quality: 166997 bases at least Q20
Insert size: 172000; agarose-1p
Insert size: 167354; sum-of-contigs
Quality coverage: 11.1 in Q20 bases; agarose-1p
Quality coverage: 11.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 20209: contig of 20209 bp in length
* 20210 20309: gap of 100 bp
* 20310 21148: contig of 839 bp in length
* 21149 21248: gap of 100 bp
* 21249 21841: contig of 593 bp in length
* 21842 21941: gap of 100 bp
* 21942 22950: contig of 1009 bp in length
* 22951 23050: gap of 100 bp
* 23051 24345: contig of 1295 bp in length
* 24346 24445: gap of 100 bp
* 24446 25663: contig of 1218 bp in length
* 25664 32382: contig of 6619 bp in length
* 32383 32483: gap of 100 bp
* 32483 52046: contig of 19564 bp in length
* 52047 52146: gap of 100 bp
* 52147 121530: contig of 69384 bp in length
* 121531 121630: gap of 100 bp
* 121631 155705: contig of 34075 bp in length
* 155706 155805: gap of 100 bp
* 155806 168354: contig of 12549 bp in length.
*
* Location/Qualifiers
* 1..168354
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
FEATURES
source

```



```
repeat_region /rpt_family="L1"
9564. .9667
/rpt_family="B4"
repeat_region /rpt_family="B4"
9694. .9985
/rpt_family="L1"
repeat_region /rpt_family="L1"
10125. .10335
/rpt_family="L1"
repeat_region /rpt_family="L1"
11552. .11666
/rpt_family="B4"
repeat_region /rpt_family="B4"
12289. .12334
/rpt_family="L1"
repeat_region /rpt_family="L1"
13538. .13598
/rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
16456. .16593
/rpt_family="B4"
repeat_region /rpt_family="B4"
16748. .17136
/rpt_family="ERVK"
repeat_region /rpt_family="ERVK"
17590. .17758
/rpt_family="L1"
repeat_region /rpt_family="L1"
18022. .18100
/rpt_family="ERVK"
repeat_region /rpt_family="ERVK"
18400. .18525
/rpt_family="RMR15"
repeat_region /rpt_family="RMR15"
18527. .18816
/rpt_family="L1"
repeat_region /rpt_family="L1"
18828. .19159
/notes="Unresolved simple sequence repeat."
repeat_region /rpt_family="RMR15"
19147. .19464
/rpt_family="RMR15"
repeat_region /rpt_family="RMR15"
20919. .21384
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
21761. .21828
/rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
22079. .22305
/rpt_family="B4"
repeat_region /rpt_family="B4"
22306. .22440
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
22441. .22514
/rpt_family="B4"
repeat_region /rpt_family="B4"
22560. .22646
/rpt_family="ERVK"
repeat_region /rpt_family="ERVK"
23695. .23772
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
24242. .24761
/rpt_family="RMR6A"
repeat_region /rpt_family="RMR6A"
24936. .26057
/rpt_family="L1"
repeat_region /rpt_family="L1"
25976. .26056
/notes="Unresolved simple sequence repeat."
repeat_region /rpt_family="MaLR"
29547. .29893
/rpt_family="MaLR"
repeat_region /rpt_family="B4"
30811. .30935
/rpt_family="B4"
repeat_region /rpt_family="B4"
33293. .33607
/rpt_family="MER2_type"
repeat_region /rpt_family="MER2_type"
33770. .34174
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
34354. .34450
/rpt_family="L2"
repeat_region /rpt_family="L2"
37257. .37539
/notes="Unresolved simple sequence repeat."
repeat_region /rpt_family="L1"
39204. .39862
/rpt_family="L1"
repeat_region /rpt_family="L1"
40201. .40267
/rpt_family="ID"
repeat_region /rpt_family="ID"
41021. .41418
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
42084. .42461
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
42766. .42907
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
42923. .43109
/rpt_family="B2"

repeat_region 43782. .43830
/rpt_family="L2"
repeat_region 44077. .44158
/rpt_family="L1"
repeat_region 45787. .45980
/rpt_family="B2"
repeat_region 46247. .46447
/rpt_family="MaLR"
repeat_region 46604. .47002
/rpt_family="MaLR"
repeat_region 47907. .48071
/rpt_family="MaLR"
repeat_region 48783. .49214
/rpt_family="L1"
repeat_region 54485. .54678
/rpt_family="L2"
repeat_region 56157. .56351
/rpt_family="L1"
repeat_region 56444. .56688
/rpt_family="L2"
repeat_region 58240. .58395
/rpt_family="MaLR"
repeat_region 58416. .58578
/rpt_family="B2"
repeat_region 58648. .59082
/rpt_family="L1"
repeat_region 65037. .65180
/rpt_family="Alu"
repeat_region 65593. .65943
/rpt_family="MaLR"
repeat_region 66012. .66310
/rpt_family="MaLR"
repeat_region 66722. .67335
/rpt_family="MaLR"
repeat_region 67336. .67681
/rpt_family="MaLR"
repeat_region 67694. .67825
/rpt_family="Alu"
repeat_region 68855. .69037
/rpt_family="B2"
repeat_region 68890. .69046
/rpt_family="B4"

Query Match 83.2%; Score 20.8; DB 9; Length 170429;
Best Local Similarity 91.7%; Pred.No.61;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGCTG 24
Db 19253 AAGAAACAATGACCACTGGCTG 19276

RESULT 9
AL590454 171970 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-266010, 15 unordered pieces.
DEFINITION
ACCESSION AL590454
VERSION AL590454.9 GI:14270736
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1
AUTHORS Mclay, K.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 31, 2001 this sequence version replaced gi:14268249.
----- Genome Center
Center: Sanger Centre
```

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: hunquary@sanger.ac.uk
----- Project Information
Center project name: BA286010
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
Dye-terminator Big Dye; 96% of reads
Consensus quality: 168624 bases at least Q40
Consensus quality: 163993 bases at least Q30
Consensus quality: 170059 bases at least Q20
Insert size: 170570; sum-of-contigs
Insert size: 187521; 2.8% error; agarose-1p
Quality coverage: 8.38x in Q20 bases; sum-of-contigs Quality
coverage: 7.68x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 6149: contig of 6149 bp in length
6150 6249: gap of 100 bp
6250 12238: contig of 5989 bp in length
12239 12338: gap of 100 bp
12339 22361: contig of 10023 bp in length
22362 22461: gap of 100 bp
22462 35419: contig of 12958 bp in length
35420 35519: gap of 100 bp
35520 41347: contig of 5828 bp in length
41348 41447: gap of 100 bp
41448 44902: contig of 3455 bp in length
44903 45002: gap of 100 bp
45003 68731: contig of 23728 bp in length
68732 68830: gap of 100 bp
68831 81218: contig of 12388 bp in length
81219 81318: gap of 100 bp
81319 104441: contig of 23123 bp in length
104442 104541: gap of 100 bp
104542 109172: contig of 4631 bp in length
109173 109272: gap of 100 bp
109273 118630: contig of 9358 bp in length
118631 118730: gap of 100 bp
118731 137728: contig of 18998 bp in length
137729 137828: gap of 100 bp
137829 145692: contig of 7864 bp in length
145693 145792: gap of 100 bp
145793 154640: contig of 8848 bp in length
154641 154740: gap of 100 bp
154741 171970: contig of 17230 bp in length.

FEATURES
    source
        1..171970
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone_lib="RP11-11.1"
            /clone="RP11-266010"
            1..6149
                /note="assembly fragment:03712"
                fragment_chain:1
                    clone_end:SP6
                    vector_side:left
                    6250..12238
                        /note="assembly fragment:02922"
                        fragment_chain:1
                            12339..22361
                                /note="assembly fragment:01816

```

```

fragment_chain:1"
22462..35419
/note="assembly fragment:00457
fragment_chain:2"
35520..41347
/note="assembly fragment:03865
fragment_chain:2"
41448..44902
/note="assembly fragment:02285
fragment_chain:3"
45003..68730
/note="assembly fragment:03722
fragment_chain:3"
68831..81218
/note="assembly fragment:01982
fragment_chain:3"
81319..104441
/note="assembly fragment:02876
fragment_chain:3"
104542..109172
/note="assembly fragment:03156
fragment_chain:3"
109273..118630
/note="assembly fragment:00850
fragment_chain:3"
118731..137728
/note="assembly fragment:02424
fragment_chain:3"
137829..145692
/note="assembly fragment:04254
fragment_chain:3"
145793..154640
/note="assembly fragment:03799
fragment_chain:3"
154741..171970
/note="assembly fragment:00365
fragment_chain:3
clone_end:T7
vector_side:right"

ORIGIN

Query Match      83.2%; Score 20.8; DB 14; Length 171970;
Best Local Similarity 91.7%; Pred.No.61;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCAACTGGGCTG 24
    ||||| ||||| ||||| ||||| |||||
Db 4799 AAAAAAATGACCAACTGGGCTG 4822

RESULT 10
AC025988 177654 bp DNA linear HTG 10-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-270L22 map 1, WORKING DRAFT
DEFINITION
SEQUENCE, 9 unordered pieces.
AC025988
AC025988 3 GI:10047868
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Homiidae; Homo.
REFERENCE
1. (bases 1 to 177654)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
JOURNAL Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
REFERENCE Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
AUTHORS Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,

```









Tel: 6172521477

Fax: 6172580903

Email: kersli@genome.wi.mit.edu

Primer A: No sequence submitted

Primer B: No sequence submitted

STS size: 601

Protocol:

WGS-discovery (WGS):

paired-end and low-coverage whole genome shotgun reads were generated from 9 breeds (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each) and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as STSs and 45941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

WGA-discovery (WGA) of Boxer/Poodle SNPs: A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the boxer and the poodle. The validation rate for these SNPs was estimated at approximately TBD%.

Internal-WGA-discovery (I-WGA): A third set of SNPs were discovered by comparing reads in the WGA assembly. SNPs were defined as mismatch positions that had a base quality of  $\geq 30$  on both reads in a region that aligned without gaps, and with at most one additional mismatch in the ten flanking bases. For each allele, at least one additional read had to confirm it. 731476 SNPs were annotated with alleles between the two boxer alleles. The validation rate for these SNPs was estimated at approximately TBD%.

```

FEATURES
  source
    1. .601
      /organism="Canis familiaris"
      /mol_type="genomic DNA"
      /strain="Beagle"
      /db_xref="taxon:9615"
      /map="9 8-534 44672481-44671955"
      /clone_lib="Beagle"
      <1. .>601

```

## ORIGIN

```

Query Match      80.8%; Score 20.2; DB 10; Length 601;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 AAAAAAATGACCAACTGGGCTGT 25

|||||

Db 201 AAAACACAGTGACCACTGGGCAGT 177

|||||

## RESULT 14

AC165059/c

LOCUS

AC165059

40745 bp

DNA

linear

HTG 02-JUL-2005

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Phakopsora pachyrhizi clone JG1AFNA-1965M23, WORKING DRAFT SEQUENCE, 2 unordered pieces.

AC165059

AC165059.1 GI:68533343

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

Phakopsora pachyrhizi

Phakopsora pachyrhizi

Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.

1 (bases 1 to 40745)

DOE Joint Genome Institute.

Unpublished

2 (bases 1 to 40745)

DOE Joint Genome Institute.

Direct Submission

Submitted (02-JUL-2005) Production Genomics Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA

94598-1698, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

-----

Project Information

Center Project Name: 4000932

Center clone name: JGI-AFNA\_1965M23

-----

Summary Statistics

Consensus quality: 40507 bases at least Q40

Consensus quality: 40595 bases at least Q30

Consensus quality: 40641 bases at least Q20

Estimated insert size: 46000; pulse field gel estimation

Estimated insert size: 40645; sum-of-contigs estimation

Quality coverage: 17.22 in Q20 bases; pulse field gel estimation

Quality coverage: 19.49 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1463: contig of 1463 bp in length

\* 1464 1563: gap of unknown length

\* 1564 40745: contig of 39182 bp in length.

## FEATURES

## source

```

    1. .40745
      /location/Qualifiers
      /organism="Phakopsora pachyrhizi"
      /mol_type="genomic DNA"
      /db_xref="taxon:170000"
      /clone="JG1AFNA-1965M23"
      /clone_lib="JGI Fosmid library AFNA"
      1464. .1563
      /estimated_length=unknown

```

## ORIGIN

```

Query Match      80.8%; Score 20.2; DB 14; Length 40745;
Best Local Similarity 88.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 AAAAAAATGACCAACTGGGCTGT 25

|||||

Db 36386 AAAAAAATGACCAACTGGGTTT 36362

|||||

## RESULT 15

AC100243

LOCUS

DEFINITION

AC100243

ACCESSION

AC100243.1

GI:17047609

5175 bp

DNA

linear

HTG 22-NOV-2001

MUS musculus clone RP23-64K4, LOW-PASS SEQUENCE SAMPLING.

KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 57175)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP23-64K4  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 57175)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepeil, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G.,  
MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, M., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L14443  
Center clone name: 64\_K\_4  
-----  
\* NOTE: This record contains 71 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 709: contig of 709 bp in length  
\* 810 809: gap of 100 bp  
\* 1521: contig of 712 bp in length  
\* 1522 1621: gap of 100 bp  
\* 1622 2345: contig of 724 bp in length  
\* 2346 2445: gap of 100 bp  
\* 2446 3149: contig of 704 bp in length  
\* 3150 3249: gap of 100 bp  
\* 3250 3954: contig of 705 bp in length  
\* 3955 4054: gap of 100 bp  
\* 4055 4768: contig of 714 bp in length  
\* 4769 4868: gap of 100 bp  
\* 4869 5577: contig of 709 bp in length  
\* 5578 5677: gap of 100 bp  
\* 6391: contig of 714 bp in length  
\* 6491: gap of 100 bp  
\* 6492 7198: contig of 707 bp in length  
\* 7199 7298: gap of 100 bp  
\* 7299 8014: contig of 716 bp in length  
\* 8015 8115: gap of 100 bp  
\* 8116 8832: contig of 718 bp in length  
\* 8833 9488: contig of 716 bp in length  
\* 9489 9749: gap of 100 bp  
\* 9750 10452: contig of 704 bp in length  
\* 10453 10552: gap of 100 bp  
\* 10553 11251: contig of 698 bp in length  
\* 11252 11350: gap of 100 bp  
\* 11351 12039: contig of 689 bp in length  
\* 12040 12139: gap of 100 bp  
\* 12140 12851: contig of 712 bp in length  
\* 12852 12951: gap of 100 bp  
\* 12952 13666: contig of 715 bp in length  
\* 13667 13766: gap of 100 bp  
\* 13767 14463: contig of 697 bp in length  
\* 14464 14563: gap of 100 bp  
\* 14564 15278: contig of 715 bp in length  
\* 15279 15378: gap of 100 bp  
\* 15379 16100: contig of 722 bp in length  
\* 16101 16200: gap of 100 bp  
\* 16201 16920: contig of 720 bp in length  
\* 16921 17020: gap of 100 bp  
\* 17021 17119: contig of 699 bp in length  
\* 17120 17819: gap of 100 bp  
\* 17820 18523: contig of 704 bp in length  
\* 18524 18623: gap of 100 bp  
\* 18624 19335: contig of 712 bp in length  
\* 19336 19435: gap of 100 bp  
\* 19436 20251: contig of 716 bp in length  
\* 20252 20961: contig of 710 bp in length  
\* 20962 21061: gap of 100 bp  
\* 21062 21776: contig of 715 bp in length  
\* 21777 21876: gap of 100 bp  
\* 21877 22600: contig of 724 bp in length  
\* 22601 22700: gap of 100 bp  
\* 22701 23421: contig of 721 bp in length  
\* 23422 23521: gap of 100 bp  
\* 23522 24328: contig of 707 bp in length  
\* 24329 25020: contig of 692 bp in length  
\* 25021 25120: gap of 100 bp  
\* 25121 25810: contig of 690 bp in length  
\* 25811 25910: gap of 100 bp  
\* 25912 26817: contig of 707 bp in length  
\* 26818 26717: gap of 100 bp  
\* 26718 27426: contig of 709 bp in length  
\* 27427 27526: gap of 100 bp  
\* 27527 28243: contig of 717 bp in length  
\* 28244 28343: gap of 100 bp  
\* 28344 29019: contig of 676 bp in length  
\* 29020 29119: gap of 100 bp  
\* 29120 29816: contig of 697 bp in length  
\* 29817 29916: gap of 100 bp  
\* 29918 30625: contig of 709 bp in length  
\* 30626 30725: gap of 100 bp  
\* 30726 31437: contig of 711 bp in length  
\* 31438 31536: gap of 100 bp  
\* 31537 32241: contig of 705 bp in length  
\* 32242 32341: gap of 100 bp  
\* 32342 33056: contig of 715 bp in length  
\* 33057 33156: gap of 100 bp  
\* 33157 33864: contig of 708 bp in length  
\* 33865 33964: gap of 100 bp  
\* 33965 34671: contig of 707 bp in length  
\* 34672 34771: gap of 100 bp  
\* 34772 35481: contig of 710 bp in length

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 62356)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 17, clone RP11-1109H12  
Unpublished

2 (bases 1 to 62356)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L25448

Center clone name: 1109\_H\_12

-----

\* NOTE: This record contains 79 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 699: contig of 699 bp in length  
\* 700: gap of 100 bp  
\* 1497: contig of 698 bp in length  
\* 1498: gap of 100 bp  
\* 1598: contig of 703 bp in length  
\* 2301: gap of 100 bp  
\* 2401: contig of 685 bp in length  
\* 3086: gap of 100 bp  
\* 3186: contig of 683 bp in length  
\* 3869: gap of 100 bp  
\* 3969: contig of 684 bp in length  
\* 4653: gap of 100 bp  
\* 4753: contig of 709 bp in length  
\* 5462: gap of 100 bp  
\* 5562: contig of 710 bp in length  
\* 6272: gap of 100 bp  
\* 7067: contig of 696 bp in length  
\* 7168: gap of 100 bp  
\* 7852: contig of 685 bp in length  
\* 7853: gap of 100 bp

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 62356)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 17, clone RP11-1109H12  
Unpublished

2 (bases 1 to 62356)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L25448

Center clone name: 1109\_H\_12

-----

\* NOTE: This record contains 79 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 699: contig of 699 bp in length  
\* 700: gap of 100 bp  
\* 1497: contig of 698 bp in length  
\* 1498: gap of 100 bp  
\* 1598: contig of 703 bp in length  
\* 2301: gap of 100 bp  
\* 2401: contig of 685 bp in length  
\* 3086: gap of 100 bp  
\* 3186: contig of 683 bp in length  
\* 3869: gap of 100 bp  
\* 3969: contig of 684 bp in length  
\* 4653: gap of 100 bp  
\* 4753: contig of 709 bp in length  
\* 5462: gap of 100 bp  
\* 5562: contig of 710 bp in length  
\* 6272: gap of 100 bp  
\* 7067: contig of 696 bp in length  
\* 7168: gap of 100 bp  
\* 7852: contig of 685 bp in length  
\* 7853: gap of 100 bp

35482 35581: gap of 100 bp  
\* 35582: contig of 705 bp in length  
\* 36287 36386: gap of 100 bp  
\* 36387: contig of 700 bp in length  
\* 37087 37186: gap of 100 bp  
\* 37187: contig of 713 bp in length  
\* 37900 37999: gap of 100 bp  
\* 38000: contig of 714 bp in length  
\* 38714 38813: gap of 100 bp  
\* 38814: contig of 692 bp in length  
\* 39506 39605: gap of 100 bp  
\* 39606: contig of 701 bp in length  
\* 40307 40406: gap of 100 bp  
\* 40407: contig of 706 bp in length  
\* 41113 41212: gap of 100 bp  
\* 41213: contig of 705 bp in length  
\* 41917 42016: gap of 100 bp  
\* 42018: contig of 707 bp in length  
\* 42725 42824: gap of 100 bp  
\* 42825: contig of 715 bp in length  
\* 43540 43639: gap of 100 bp  
\* 43640: contig of 715 bp in length  
\* 44355 44454: gap of 100 bp  
\* 44455: contig of 681 bp in length  
\* 45136 45235: gap of 100 bp  
\* 45236: contig of 712 bp in length  
\* 45948 46047: gap of 100 bp  
\* 46048: contig of 701 bp in length  
\* 46749 46848: gap of 100 bp  
\* 46849: contig of 704 bp in length  
\* 47553 47652: gap of 100 bp  
\* 47653: contig of 699 bp in length  
\* 48352 48451: gap of 100 bp  
\* 48452: contig of 671 bp in length  
\* 49123 49222: gap of 100 bp  
\* 49223: contig of 713 bp in length  
\* 49936 50035: gap of 100 bp  
\* 50036: contig of 712 bp in length  
\* 50748 50847: gap of 100 bp  
\* 50848: contig of 673 bp in length  
\* 51521 51620: gap of 100 bp  
\* 51621: contig of 716 bp in length  
\* 52337 52436: gap of 100 bp  
\* 52437: contig of 701 bp in length  
\* 53138 53237: gap of 100 bp  
\* 53238: contig of 716 bp in length  
\* 53954 54053: gap of 100 bp  
\* 53954: contig of 703 bp in length  
\* 54054 54153: gap of 100 bp  
\* 54154: contig of 703 bp in length  
\* 54757 54856: gap of 100 bp  
\* 54857: contig of 707 bp in length

Query Match 80.8%; Score 20.2; DB 14; Length 57175;  
Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 AAAAAACATGACCACTGGCGTGT 25  
Db 20553 AAGAAACAAAGACCACTGAGCTGT 20577

RESULT 16  
AC111185  
LOCUS AC111185 62356 bp DNA linear HTG 18-FEB-2002  
DEFINITION Homo sapiens chromosome 17 clone RP11-1109H12 map 17, LOW-PASS  
SEQUENCE SAMPLING.  
AC111185  
VERSION AC111185.1 GI:186999950  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

```
* 7953 8667: contig of 715 bp in length
* 8668 8767: gap of 100 bp
* 8768 9471: contig of 704 bp in length
* 9472 9571: gap of 100 bp
* 9572 10278: contig of 707 bp in length
* 10279 11060: gap of 100 bp
* 11061 11160: contig of 682 bp in length
* 11161 11856: gap of 100 bp
* 11857 11856: contig of 696 bp in length
* 11857 11956: gap of 100 bp
* 11957 12635: contig of 679 bp in length
* 12636 12735: gap of 100 bp
* 12736 13444: contig of 709 bp in length
* 13445 13544: gap of 100 bp
* 13545 14240: contig of 696 bp in length
* 14241 14340: gap of 100 bp
* 14341 15010: contig of 670 bp in length
* 15011 15110: gap of 100 bp
* 15111 15799: contig of 689 bp in length
* 15800 15899: gap of 100 bp
* 15900 16562: contig of 663 bp in length
* 16563 16662: gap of 100 bp
* 16664 17358: contig of 696 bp in length
* 17359 17458: gap of 100 bp
* 17459 18164: contig of 706 bp in length
* 18165 18264: gap of 100 bp
* 18265 18945: contig of 681 bp in length
* 18946 19045: gap of 100 bp
* 19046 19726: contig of 681 bp in length
* 19727 19826: gap of 100 bp
* 19827 20515: contig of 689 bp in length
* 20516 20615: gap of 100 bp
* 20616 21326: contig of 711 bp in length
* 21327 21426: gap of 100 bp
* 21427 22139: contig of 713 bp in length
* 22140 22239: gap of 100 bp
* 22240 22925: contig of 686 bp in length
* 22926 23025: gap of 100 bp
* 23026 23703: contig of 678 bp in length
* 23704 23803: gap of 100 bp
* 23804 24507: contig of 704 bp in length
* 24508 24607: gap of 100 bp
* 24608 25308: contig of 701 bp in length
* 25309 25408: gap of 100 bp
* 25409 26075: contig of 667 bp in length
* 26076 26175: gap of 100 bp
* 26176 26863: contig of 688 bp in length
* 26864 26963: gap of 100 bp
* 26964 27552: contig of 689 bp in length
* 27553 27653: gap of 100 bp
* 27653 28466: contig of 714 bp in length
* 28467 28566: gap of 100 bp
* 28567 29273: contig of 707 bp in length
* 29274 29373: gap of 100 bp
* 29374 30050: contig of 677 bp in length
* 30051 30150: gap of 100 bp
* 30151 30831: contig of 681 bp in length
* 30832 30931: gap of 100 bp
* 30932 31592: contig of 661 bp in length
* 31593 31692: gap of 100 bp
* 31693 32381: contig of 689 bp in length
* 32382 32481: gap of 100 bp
* 32482 33189: contig of 708 bp in length
* 33190 33289: gap of 100 bp
* 33290 33983: contig of 694 bp in length
* 33984 34084: gap of 100 bp
* 34084 34773: contig of 690 bp in length
* 34774 34873: gap of 100 bp
* 34874 35551: contig of 678 bp in length
* 35552 36358: gap of 100 bp
* 36359 36458: contig of 707 bp in length
* 36459 37150: gap of 100 bp
* 37150 37150: contig of 692 bp in length
```

```
* 37151 37250: gap of 100 bp
* 37251 37927: contig of 677 bp in length
* 37928 38027: gap of 100 bp
* 38028 38685: contig of 658 bp in length
* 38686 38785: gap of 100 bp
* 38786 39489: contig of 704 bp in length
* 39490 39589: gap of 100 bp
* 39590 40287: contig of 698 bp in length
* 40288 40387: gap of 100 bp
* 40388 41080: contig of 693 bp in length
* 41081 41180: gap of 100 bp
* 41181 41863: contig of 683 bp in length
* 41864 41963: gap of 100 bp
* 41964 42642: contig of 679 bp in length
* 42643 42742: gap of 100 bp
* 42743 43415: contig of 673 bp in length
* 43416 43515: gap of 100 bp
* 43516 44199: contig of 684 bp in length
* 44200 44299: gap of 100 bp
* 44300 44993: contig of 694 bp in length
* 44994 45093: gap of 100 bp
* 45094 45785: contig of 692 bp in length
* 45786 45885: gap of 100 bp
* 45886 46565: contig of 680 bp in length
* 46566 47360: gap of 100 bp
* 47361 47460: contig of 695 bp in length
* 47461 48140: contig of 680 bp in length
* 48141 48240: gap of 100 bp
* 48241 48922: contig of 682 bp in length
* 48923 49022: gap of 100 bp
* 49023 49711: contig of 689 bp in length
* 49712 49811: gap of 100 bp
* 49812 50496: contig of 685 bp in length
* 50497 50596: gap of 100 bp
* 50597 51294: contig of 698 bp in length
* 51295 51394: gap of 100 bp
* 51395 52092: contig of 698 bp in length
* 52093 52192: gap of 100 bp
* 52193 52905: contig of 713 bp in length
* 52906 53005: gap of 100 bp
* 53006 53703: contig of 698 bp in length
* 53704 53803: gap of 100 bp
```

Query Match 80.8%; Score 20.2; DB 14; Length 62356;  
Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCAACTGGGCTGT 25

Db 21910 AAAAAACATGACCAACTGGGCTGT 21934

RESULT 17  
AL356141/c

LOCUS Human DNA sequence from clone Rp11-269L3 on chromosome 10, complete  
DEFINITION sequence.

ACCESSION AL356141  
VERSION AL356141.9 GI:15617255

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE 1 (bases 1 to 87871)  
AUTHORS Heath, P.  
TITLE Direct Submission

JOURNAL

COMMENT Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Sep 13, 2001 this sequence version replaced gi:14970334.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
 RP11-269L3 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see  
<http://www.choi.org/bacpac/home.htm>  
 VECTOR: pBACe3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

#### FEATURES

source  
 1. 87871  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-269L3"  
 /clone\_lib="RP11-11.1"  
 /note="Clone\_left\_end: RP11-269L3"  
 85872  
 /note="Clone\_left\_end: RP11-47B24"

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 87871;  
 Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 AAAAAACAATGACCACTGGCTGT 25  
 Db 6325 AAAAAACAATGACCACTGGCTGT 6301

#### RESULT 18

AC154170  
 LOCUS  
 DEFINITION Alligator mississippiensis clone VMRC8-371122, WORKING DRAFT  
 AC154170 134231 bp DNA linear HTG 24-FEB-2005  
 SEQUENCE, 8 ordered pieces.  
 AC154170  
 AC154170.2 GI:60223218  
 HTG; HTGS PHASE2; HTGS DRAFT.  
 Alligator mississippiensis (American alligator)  
 Alligator mississippiensis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Crocodylidae; Alligatorinae; Alligator.

#### REFERENCE

AUTHORS  
 1 (bases 1 to 134231)  
 Antonellis,A., Ayale,K., Benjamin,B., Blakesley,R.W., Boake,A.,  
 Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,  
 Engle,J., Gestole,M., Guan,X., Gupta,J., Gutierrez,P., Haghighi,P.,  
 Han,J., Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurle,B., Idol,J.R.,  
 Jones,C., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R.,  
 Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,  
 Maskeri,B., McDowell,J., Mullikin,J.C., Oestreich,J.S., Park,M.,  
 Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N., Rosas,B.,  
 Schandler,K., Schueler,M.G., Sison,C., Stantropop,S., Stephen,E.,  
 Tave,A., Thomas,J.W., Thomas P.J., Tsipouri,V., Vogt,J.L.,  
 Wetherby,K.D., Young,A. and Green,E.D.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 134231)  
 Green,E.D.  
 Direct Submission  
 Submitted (28-DEC-2004) NIH Intramural Sequencing Center, 5625  
 Fishers Lane, Rockville, MD 20852, USA  
 3 (bases 1 to 134231)  
 Green,E.D.  
 Direct Submission  
 Submitted (24-FEB-2005) NIH Intramural Sequencing Center, 5625  
 Fishers Lane, Rockville, MD 20852, USA  
 On Feb 24, 2005 this sequence version replaced gi:56806706.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc.zoo@nhgri.nih.gov](mailto:nisc.zoo@nhgri.nih.gov)  
 ----- Project Information  
 Center project name: gnt  
 Center clone name: 371122

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig, has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is generally based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

#### ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 133035 bases at least Q40  
 Consensus quality: 133354 bases at least Q30  
 Consensus quality: 133486 bases at least Q20  
 Insert size: 152000; agarose-1p  
 Insert size: 133531; sum-of-contigs  
 Quality coverage: 10.18x in Q20 bases; agarose-fp  
 Quality coverage: 11.59x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 26879: contig of 26879 bp in length  
 \* 26880 26979: gap of unknown length  
 \* 26980 29523: contig of 2544 bp in length  
 \* 29524 29623: gap of unknown length  
 \* 29624 67124: contig of 37401 bp in length  
 \* 67025 67124: gap of unknown length  
 \* 67125 92702: contig of 25578 bp in length  
 \* 92703 92802: gap of unknown length  
 \* 92803 95901: contig of 3099 bp in length  
 \* 95902 96001: gap of unknown length  
 \* 96002 100291: contig of 4290 bp in length  
 \* 100292 100391: gap of unknown length  
 \* 100392 129050: contig of 28659 bp in length  
 \* 129051 129150: gap of unknown length  
 \* 129151 134231: contig of 5081 bp in length.  
 Location/Qualifiers  
 1. 134231  
 /organisms="Alligator mississippiensis"

#### FEATURES

source



misc\_feature 1. 35894 /mol\_type="genomic DNA" /db\_xref="taxon:8496" /clone="VMRC8-371122" /clone\_lib="VMRC8" /note="BAC resource: http://bacpac.chori.org/"

misc\_feature AC154169 clone overlaps with GenBank Accession Number AC154169 clone VMRC8-246114 (center project name gns) 1. .26879 /note="assembly\_fragment clone\_end:r7 vector\_side:left" 26880. .26979 /estimated\_length=unknown 26980. .29523 /note="assembly\_fragment" 29524. .29623 /estimated\_length=unknown 29624. .67024 /notes="assembly\_fragment" 67025. .67124 /estimated\_length=unknown 67125. .92702 /notes="assembly\_fragment" 92703. .92802 /estimated\_length=unknown 92803. .95901 /note="assembly\_fragment" 95902. .96001 /estimated\_length=unknown 96002. .100291 /notes="assembly\_fragment" 100292. .100391 /estimated\_length=unknown 100392. .129050 /notes="assembly\_fragment" 129051. .129150 /estimated\_length=unknown 129151. .134231 /notes="assembly\_fragment clone\_end:SP6 vector\_side:right"

ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 134231; Best Local Similarity 88.8%; Pred. No. 1.2e+02; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCACTGGGCTGT 25  
||||| ||||| ||||| ||||| |||||

Db 52652 AAAAAAGAACCAATCGGCTGT 52676

RESULT 19

AC090735

LOCUS AC090735 149089 bp DNA linear PRI 20-FEB-2002

DEFINITION Homo sapiens chromosome 8, clone RP11-1145L24, complete sequence.

ACCESSION AC090735

VERSION AC090735.5 GI:18767535

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 149089)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 8, clone RP11-1145L24

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 149089)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Fereira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Laroque,K., Lamazares,R., Landers,T., Lehotsky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L12821

Center clone name: 1145\_L\_24

Location/Qualifiers

1. 149089  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-1145L24"  
/clone\_lib="RP11-1145L24"  
complement(497..603)  
/rpt\_family="L2"  
839..885  
/rpt\_family="L2"

repeat\_region  
repeat\_region

```

repeat_region 1542. .1579
/rpt family="CT-rich"
complement(1616. .1924)
/rpt family="ALUSC"
repeat_region complement(4056. .4126)
/rpt family="L2"
repeat_region 4128. .4191
/rpt family="GA-rich"
repeat_region 4293. .4344
/rpt family="L1PA6"
complement(4458. .4635)
/rpt family="L1PA10"
repeat_region complement(5395. .5571)
/rpt family="FRAM"
repeat_region 5588. .5848
/rpt family="L1R16C"
complement(6035. .6096)
/rpt family="MIR"
repeat_region 6456. .7176
/rpt family="L2"
repeat_region 7182. .7569
/rpt family="L1MC4a"
complement(7927. .8978)
/rpt family="L1P4"
repeat_region complement(8979. .9113)
/rpt family="FLAM C"
repeat_region complement(9114. .9948)
/rpt family="L1P4"
repeat_region 9949. .10342
/rpt family="L1PA16"
complement(10545. .10779)
/rpt family="L1PA5"
repeat_region complement(13384. .13882)
/rpt family="MLT1D"
repeat_region 14543. .15079
/rpt family="L2"
repeat_region 15694. .15714
/rpt family="AT-rich"
repeat_region 16311. .16425
/rpt family="MIR"
repeat_region 18284. .18331
/rpt family="MIR"
repeat_region 18955. .19223
/rpt family="ALUSx"
complement(19282. .19300)
/rpt family="L2"
repeat_region 19282. .19300
complement(19294)
/rpt family="L2"
repeat_region complement(19302. .19595)
/rpt family="AluJb"
repeat_region 19673. .19769
/rpt family="MIR"
repeat_region 20591. .20907
/rpt family="AluJb"
complement(22175. .22233)
/rpt family="L2"
repeat_region 23342. .23692
/rpt family="L1"
repeat_region 23704. .24007
/rpt family="L1ME"
repeat_region 24024. .24342
/rpt family="L1ME"
repeat_region 24360. .24561
/rpt family="L1ME"
repeat_region 24691. .25221
/rpt family="L1ME"
repeat_region complement(26037. .26225)
/rpt family="Charlie8"
repeat_region 26445. .26480
/rpt family="L1"
repeat_region 27413. .33443
/rpt family="L1HS"
complement(33546. .33700)

```

```

repeat_region /rpt family="MERSA"
complement(35431. .35652)
/rpt family="MIR"
repeat_region complement(35984. .36109)
/rpt family="L2"
repeat_region complement(37679. .37730)
/rpt family="L2"
repeat_region complement(37809. .37980)
/rpt family="MIR"
repeat_region complement(37990. .38222)
/rpt family="AluJb"
complement(38258. .38358)
/rpt family="MIR"
repeat_region complement(38653. .39585)
/rpt family="L1MB1"
repeat_region complement(39586. .39661)
/rpt family="ORSL"
complement(40402. .40703)
/rpt family="AluSc"
unsure 40591. .40595
/note="<30 qual SNGL region"
unsure 40634. .40716
/note="single clone coverage"
unsure 40655. .40660
/note="<30 qual single clone coverage"
unsure 40682. .40705
/note="<30 qual single clone coverage"
unsure 40709. .40716
/note="<30 qual single clone coverage"
repeat_region 40758. .40941
/rpt family="FRAM"
repeat_region complement(41497. .41634)
/rpt family="FLAM C"
repeat_region complement(41889. .42129)
/rpt family="L2"
repeat_region complement(42956. .43280)

```

Query Match 80.8%; Score 20.2; DB 8; Length 149089;  
 Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGCTGT 25

Db 91563 AAAAAATCAATGACTATCGGCTGT 91587

RESULT 20

AC012339/c

LOCUS

DEFINITION

AC012339

AC012339

AC012339.8

HTG.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC012339 179818 bp DNA linear PRI 30-DEC-2001  
 Homo sapiens chromosome 8, clone RP11-31K23, complete sequence.

AC012339  
 Homo sapiens chromosome 8, clone RP11-31K23, complete sequence.  
 AC012339  
 Homo sapiens chromosome 8, clone RP11-31K23  
 AC012339.8 GI:17998717  
 HTG.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

1 (bases 1 to 179818)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 8, clone RP11-31K23  
 Unpublished  
 2 (bases 1 to 179818)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Castle, A., Colangelo, R., Collins, S., Collymore, A.,  
 Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Lehotzky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,  
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,



Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 990315  
Consensus quality: 176000 bases at least Q40  
Consensus quality: 177396 bases at least Q30  
Consensus quality: 178472 bases at least Q20  
Insert size: 180333; sum-of-contigs  
Quality coverage: 6.1x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1524: contig of 1524 bp in length  
\* 1525: gap of unknown length  
\* 1625: contig of 32742 bp in length  
\* 34367: gap of unknown length  
\* 34467: contig of 60856 bp in length  
\* 95323: gap of unknown length  
\* 95423: contig of 85161 bp in length.  
\*  
\* Location/Qualifiers  
\* 1. .180583  
\* /organism="Homo sapiens"  
\* /mol\_type="genomic DNA"  
\* /db\_xref="taxon:9606"  
\* /chromosome="10"  
\* /clone="RP11-431P18"  
\* /clone\_lib="RPCI-11"  
\* 1. .1524  
\* /note="assembly\_name:Contig3"  
\* 1525. .1624  
\* /estimated\_length=unknown  
\* 1625. .34366  
\* /note="assembly\_name:Contig4"  
\* 34367. .34466  
\* /estimated\_length=unknown  
\* 34467. .95322  
\* /note="assembly\_name:Contig5"  
\* 95323. .95422  
\* /estimated\_length=unknown  
\* 95423. .180583  
\* /note="assembly\_name:Contig6"

FEATURES  
source  
misc\_feature  
gap  
misc\_feature  
gap  
misc\_feature  
gap  
misc\_feature

ORIGIN  
Query Match 80.8%; Score 20.2; DB 14; Length 180583;  
Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 65206 AAAAAACAATGACCACTGAGGTGT 65182  
|||||  
  
RESULT 22  
AC013562/c  
LOCUS  
DEFINITION Homo sapiens chromosome 8, clone RP11-313C15, complete sequence.  
ACCESSION AC013562  
VERSION AC013562.6 GI:13443208  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 188755)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 8, clone RP11-313C15  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 188755)

/rpt\_family="L3"  
29373. .29546  
/rpt\_family="MIR"  
complement(30916. .31155)  
/rpt\_family="MIR"  
31657. .31932  
/rpt\_family="L2"  
31999. .32168  
/rpt\_family="L2"  
33449. .33692  
/rpt\_family="MIR"  
complement(34755. .35194)  
/rpt\_family="MLT2F"  
35805. .35842  
/rpt\_family="AT\_rich"  
38261. .38368  
/rpt\_family="LRI16C"  
38413. .38432  
/rpt\_family="(TG)n"  
38601. .38626  
/rpt\_family="(TA)n"  
38633. .39406  
/rpt\_family="LLMC4a"  
39407. .39699  
/rpt\_family="AluSg"  
39700. .40505  
/rpt\_family="LLMC4a"  
complement(40516. .40992)  
/rpt\_family="MLT1D"

Query Match 88.0%; Score 20.2; DB 8; Length 179818;  
Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 67424 AACAACTAATGACCAACTGGGCTGT 67400  
|||||  
  
RESULT 21  
AC068709/c  
LOCUS  
DEFINITION AC068709 180583 bp DNA linear HTG 22-MAY-2002  
Homo sapiens chromosome 10 clone RP11-431P18, WORKING DRAFT  
SEQUENCE, 4 unordered pieces.  
ACCESSION AC068709.4 GI:14290388  
VERSION HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN;  
KEYWORDS HTGS CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 180583)  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
Sequence Data  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 180583)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAY-2000) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
COMMENT On Jun 5, 2001 this sequence version replaced gi:14150408.  
-----  
Center: Genome Therapeutics Corporation  
Center code: GTC  
Web site: <http://www.genomecorp.com/>  
Contact: [gtc-seqcenter@genomecorp.com](mailto:gtc-seqcenter@genomecorp.com)  
----- Project Information  
Center project name: hg293  
----- Summary Statistics  
Sequencing vector: N/A

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Casieff, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Doneilan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Liu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Seaman, S., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 188755)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schubback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (01-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

4 (bases 1 to 188755)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schubback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

5 (bases 1 to 188755)

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D.,

TITLE  
JOURNAL

## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## FEATURES

source

Location/Qualifiers

1. .188755

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-313C15"

/clone\_lib="RPC1-11 Human Male BAC"

117. .138

/rpt\_family="AT\_rich"

191. -213

/rpt\_family="AT\_rich"

complement(239. .866)

/rpt\_family="MER67B"

complement(1061. .1271)

/rpt\_family="AluJo"

1272. .1335

/rpt\_family="AT\_rich"

2094. .2293

/rpt\_family="L1PA16"

complement(3162. .3324)

/rpt\_family="MIR3"

3931. .3954

/rpt\_family="AT\_rich"

4511. .4829

/rpt\_family="L2"

complement(4809. .4895)

/rpt\_family="MER91B"

complement(4898. .5046)

/rpt\_family="MER91B"

5047. .5327

/rpt\_family="L2"

complement(5501. .5920)

/rpt\_family="MLT1C"

5974. .6265

/rpt\_family="L2"

6287. .6489

/rpt\_family="MER58A"

complement(6811. .6902)

/rpt\_family="MIR"

```

repeat_region 7489. .8170
/rpt_family="L1PA15"
repeat_region complement(8539. .8655)
/rpt_family="L2"
repeat_region 9751. .9786
/rpt_family="AT_rich"
repeat_region 10565. .12681
/rpt_family="Tigger1"
repeat_region complement(13712. .14008)
/rpt_family="AluSx"
repeat_region 14599. .14639
/rpt_family="(TTTTG)n"
repeat_region complement(15742. .15915)
/rpt_family="MIR"
repeat_region complement(15920. .15930)
/notes="Less than 30 qual SNGL region"
repeat_region complement(17339. .18108)
/rpt_family="L2"
repeat_region 18277. .18333
/rpt_family="L2"
repeat_region complement(19352. .19478)
/rpt_family="AluSp/q"
repeat_region 19684. .19706
/rpt_family="(TG)n"
repeat_region 19707. .19766
/rpt_family="(GA)n"
repeat_region 19767. .19923
/rpt_family="MIR"
repeat_region 19922. .20636
/rpt_family="L2"
repeat_region complement(20637. .21232)
/rpt_family="MER77"
repeat_region 21233. .21517
/rpt_family="L2"
repeat_region 21617. .21649
/rpt_family="(TG)n"
repeat_region 22651. .22807
/rpt_family="MIR3"
repeat_region 23200. .23379

```

```

Query Match 80.8%; Score 20.2; DB 8; Length 188755;
Best Local Similarity 88.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACATGACCAACTGGGCTGT 25
||||| ||||| ||||| ||||| |||||
Db 170134 AAAAAACATGACTATCTGGGCTGT 170110

```

```

RESULT 23
AC023376 189963 bp DNA linear HTG 04-JUL-2001
LOCUS Homo sapiens chromosome 8 clone RP11-687M23 map 8, WORKING DRAFT
DEFINITION SEQUENCE, 4 unordered pieces.
AC023376
AC023376.4 GI:14595907
KEYWORDS HTG; HTGS_PRAISE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 189963)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguski,M., Brown,A.,
Bouckhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenstermaker,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,

```

```

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneas,L., Mihova,T., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M.,
Norman,C.H., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triguillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced gi:7139677.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4260
Center clone name: 687_M23
----- Summary Statistics
Sequencing vector: M13; M77815; 47% of reads
Sequencing vector: plasmid; n/a; 53% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188836 bases at least Q40
Consensus quality: 189476 bases at least Q30
Consensus quality: 189595 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 189663; sum-of-contigs
Quality coverage: 10.7 in Q20 bases; agarose-fp
Quality coverage: 10.7 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 37424: contig of 37424 bp in length
* 37425 37524: gap of 100 bp
* 37525 42149: contig of 4625 bp in length
* 42150 42249: gap of 100 bp
* 42250 177306: contig of 135057 bp in length
* 177307 177406: gap of 100 bp
* 177407 189963: contig of 12557 bp in length.

FEATURES             Location/Qualifiers
     source            1. .189963
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
     misc_feature      1. .37424
                     /clone="RP11-687M23"
                     /clone_lib="RP11-11 Human Male BAC"
     misc_feature      1. .37424
                     /notes="assembly_fragment"
                     clone_end:SP6
                     vector_side:left
                     37425. .37524
                     /estimated_length=100
                     37525. .42149

```

```
/note="assembly_fragment"
42150..42249
/estimated_length=100
misc_feature
42250..177306
/note="assembly_fragment"
177307..177406
/estimated_length=100
misc_feature
177407..189963
/note="assembly_fragment
clone_end:r7
vector_side:right"

ORIGIN
Query Match      80.8%; Score 20.2; DB 14; Length 189963;
Best Local Similarity 88.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCACTGGGCTGT 25
    ||| ||||| ||||| |||||
Db 72097 AACAACTAATGACCACTGGGCTGT 72121

RESULT 24
AC074145/c      208161 bp DNA linear HTG 15-JUL-2000
LOCUS           Mus musculus clone RP23-104111, WORKING DRAFT SEQUENCE, 29
DEFINITION      unordered pieces.
ACCESSION       AC074145
VERSION         AC074145.1 GI:9211255
KEYWORDS        HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS         DOE Joint Genome Institute.
TITLE           Sequencing of Mouse
JOURNAL         Unpublished
REFERENCE       2 (bases 1 to 208161)
AUTHORS         DOE Joint Genome Institute.
TITLE           Direct Submission
JOURNAL         Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
                Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----

Project Information
Center Project Name: 0
Center Clone Name: RPCI-23_104111
-----

Summary Statistics
Consensus quality: 144587 bases at least Q40
Consensus quality: 166301 bases at least Q30
Consensus quality: 177710 bases at least Q20
Estimated insert size: 203300; agarose-fp estimation
Estimated insert size: 205361; sum-of-contigs estimation
Quality coverage: 3.24 in Q20 bases; agarose-fp estimation
Quality coverage: 3.2 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1559: contig of 1559 bp in length
* 1560 1659: gap of unknown length
* 1660 3192: contig of 1533 bp in length
* 3193 3292: gap of unknown length
```

```

3293 4690: contig of 1398 bp in length
4691 4790: gap of unknown length
4791 6321: contig of 1531 bp in length
6322 6421: gap of unknown length
6422 7828: contig of 1407 bp in length
7829 7928: gap of unknown length
7929 9433: contig of 1505 bp in length
9434 9534: gap of unknown length
9534 11091: contig of 1558 bp in length
11092 11191: gap of unknown length
11192 12672: contig of 1481 bp in length
12673 12772: gap of unknown length
12773 13889: contig of 1117 bp in length
13890 13990: gap of unknown length
13991 16328: contig of 2339 bp in length
16329 16428: gap of unknown length
16429 18331: contig of 1903 bp in length
18332 18431: gap of unknown length
18432 20562: contig of 2131 bp in length
20563 20662: gap of unknown length
20663 22799: contig of 2137 bp in length
22800 22899: gap of unknown length
22900 24688: contig of 1789 bp in length
24689 24788: gap of unknown length
24789 28947: contig of 4159 bp in length
28948 29047: gap of unknown length
29049 33636: contig of 4589 bp in length
33637 37008: contig of 3272 bp in length
37009 37108: gap of unknown length
37109 42485: contig of 5377 bp in length
42486 42586: gap of unknown length
42587 50240: contig of 7654 bp in length
50241 50339: gap of unknown length
50340 58128: contig of 7788 bp in length
58129 58228: gap of unknown length
58229 65777: contig of 7549 bp in length
65778 65877: gap of unknown length
65878 75167: contig of 9291 bp in length
75168 75267: gap of unknown length
75268 84181: contig of 8914 bp in length
84182 84282: gap of unknown length
84283 97293: contig of 13012 bp in length
97294 97393: gap of unknown length
97394 116122: contig of 18729 bp in length
116123 116222: gap of unknown length
116223 136122: contig of 19900 bp in length
136123 136222: gap of unknown length
136223 159176: contig of 22954 bp in length
159177 159276: gap of unknown length
159277 183073: contig of 23797 bp in length
183074 183173: gap of unknown length
183174 208161: contig of 24988 bp in length.

Location/Qualifiers
1..208161
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-104111"
/clone_lib="RPCI mouse BAC library 23"
1560..1659
/estimated_length=unknown
3193..3292
/estimated_length=unknown
4691..4790
/estimated_length=unknown
6322..6421
/estimated_length=unknown
7829..7928
/estimated_length=unknown
9434..9533
/estimated_length=unknown
11092..11191
/estimated_length=unknown

FEATURES
source
```





Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geert, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowale, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Lousegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, W., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pawkele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL

REFERENCE  
2 (bases 1 to 246315)

Unpublished

Worley, K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 246315)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 9, 2003 this sequence version replaced gi:24818682.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GAWW

Center Clone name: CH230-4p9

----- Summary Statistics

Assembly program: Atlas

Consensus quality: 197283 bases at least Q40

Consensus quality: 202460 bases at least Q30

Consensus quality: 206455 bases at least Q20

Estimated insert size: 211719; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. \* as soon as it is available and the accession number will be preserved.

\* 1 242724: contig of 242724 bp in length  
\* 242725 242824: gap of unknown length  
\* 242825 243872: contig of 1048 bp in length  
\* 243873 243972: gap of unknown length  
\* 243973 245076: contig of 1104 bp in length  
\* 245077 245176: gap of unknown length  
\* 245177 246315: contig of 1139 bp in length.

FEATURES

Location/Qualifiers  
1..246315  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-4p9"

1..1812

/note="wgs\_end\_extension"

Clone end: T7

3382..4325

/note="clone\_boundary"

Clone end: T7

site: EcoRI

end\_sequence: BH310992"

25124..27395

/note="wgs\_contig"

107600..108656

/note="wgs\_contig"

complement(241125..241757)

/note="clone\_boundary"

Clone end: Sp6

site: EcoRI

end\_sequence: BH310993"

242725..242824

/estimated\_length=unknown

243873..243972

/estimated\_length=unknown

245077..245176

/estimated\_length=unknown

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 14; Length 246315;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25

|||||

Db 56873 AAAAAACAATGACCAACTGGGCTGT 56849

RESULT 27

AC094299

LOCUS

DEFINITION

AC094299

AC094299

AC094299

AC094299

AC094299

AC094299

AC094299

AC094299

AC094299

AC094299

SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 249661) Muzny, D. Marie., Metzker, M., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Guebragregis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherz, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shivatsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
REFERENCE	2 (bases 1 to 249661)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 249661)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 9, 2003 this sequence version replaced gi:23322519. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GAIT  
Center clone name: CH230-314  
----- Summary Statistics  
Assembly program: Atlas;  
Consensus quality: 219534 bases at least Q40  
Consensus quality: 221593 bases at least Q30  
Consensus quality: 222699 bases at least Q20  
Estimated insert size: 234897; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 243876: contig of 243876 bp in length  
\* 243877 243976: gap of unknown length  
\* 243977 245492: contig of 1516 bp in length  
\* 245493 245592: gap of unknown length  
\* 245593 246913: contig of 1321 bp in length  
\* 246914 247014: gap of unknown length  
\* 247014 249661: contig of 2648 bp in length.  
----- Location/Qualifiers  
1. 249661  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-314"  
1. 2422  
/notes="wgs\_end\_extension  
clone\_end:T7"  
complement(13592..14341)  
/notes="clone\_boundary.  
clone\_end:T7  
site:ECORI  
end sequence:BH311293"  
224207..224703  
/notes="clone\_boundary  
clone\_end:Sp6  
site:ECORI  
end sequence:BH311296"  
225324..226449  
/notes="wgs\_end\_extension  
clone\_end:Sp6"  
232292..233410  
/notes="wgs\_end\_extension  
clone\_end:Sp6"  
241777..243876  
/notes="wgs\_end\_extension  
clone\_end:Sp6"  
243877..243976  
/estimated\_length=unknown  
245493..245592  
-----  
FEATURES  
source  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
gap  
gap

Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

**SOURCE INFORMATION:**  
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (<http://www.resgen.com>) or Pieter de Jong and co-workers at <http://www.bacpac.chori.org>.

**NEIGHBORING SEQUENCE INFORMATION:**  
This sequence is the entire insert of the clone. This clone is overlapped by AC140950.

**FEATURES**  
source  
location/Qualifiers  
1..160620  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="7"  
/map="7"  
/clone="RP43-31L17"  
/clone\_lib="RPCI-43"

**ORIGIN**  
Query Match 79.2%; Score 19.8; DB 8; Length 160620;  
Best Local Similarity 91.3%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAACAATGACCAACTGGGTG 24  
|||||  
Db 63994 AAAAAACAATGACCAACAGGATG 64016

**RESULT 29**  
AC166995/c  
LOCUS  
DEFINITION  
AC166995.1 GI:72535010  
VERSION  
HTG; HTGS PHASE1.  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.

**REFERENCE**  
1 (bases 1 to 170524)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 170524)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
Submitted (13-AUG-2005) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

**COMMENT**  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
----- Project Information -----  
Center project name: M\_BA0255N15  
----- Summary Statistics -----  
Sequencing vector: pL3; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 161427 bases at least Q40  
Consensus quality: 163374 bases at least Q30

gap  
/estimated\_length=unknown  
246914..247013  
/estimated\_length=unknown

**ORIGIN**  
Query Match 80.8%; Score 20.2; DB 14; Length 249661;  
Best Local Similarity 88.0%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACAATGACCAACTGGGTG 25  
|||||  
Db 15854 AAAAAACAATGACCAACTGAGCTGT 15878

**RESULT 28**  
AC146241  
LOCUS  
DEFINITION  
AC146241  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.

**REFERENCE**  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 160620)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA

**REFERENCE**  
3 (bases 1 to 160620)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
Submitted (12-MAY-2004) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA

**REFERENCE**  
4 (bases 1 to 160620)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
Submitted (19-JUN-2004) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA

**REFERENCE**  
5 (bases 1 to 160620)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
Submitted (30-JUN-2004) Washington University School of Medicine,  
Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO  
63108, USA

**COMMENT**  
On Jun 19, 2004 this sequence version replaced gi:47131419.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Summary Statistics -----  
Center project name: C\_PT031L17

**NOTICE:**  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. Wes Warren,

Consensus quality: 164740 bases at least Q20

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----

1 1251: contig of 1251 bp in length  
\* 1252: gap of unknown length  
\* 1352: contig of 1932 bp in length  
\* 3283: gap of unknown length  
\* 3284: contig of 1258 bp in length  
\* 4641: gap of unknown length  
\* 4741: gap of unknown length  
\* 8114: contig of 3373 bp in length  
\* 8215: gap of unknown length  
\* 13181: contig of 4967 bp in length  
\* 13282: contig of 2727 bp in length  
\* 16008: gap of unknown length  
\* 16109: contig of 5007 bp in length  
\* 21115: gap of unknown length  
\* 21215: contig of 3529 bp in length  
\* 21216: gap of unknown length  
\* 24745: gap of unknown length  
\* 24845: contig of 7411 bp in length  
\* 32255: gap of unknown length  
\* 32356: contig of 7871 bp in length  
\* 40227: gap of unknown length  
\* 40327: contig of 6706 bp in length  
\* 47032: gap of unknown length  
\* 47133: contig of 8288 bp in length  
\* 55421: gap of unknown length  
\* 55521: contig of 11338 bp in length  
\* 66859: gap of unknown length  
\* 66959: contig of 12357 bp in length  
\* 79315: gap of unknown length  
\* 79416: contig of 18688 bp in length  
\* 98104: gap of unknown length  
\* 98204: contig of 35227 bp in length  
\* 133431: gap of unknown length  
\* 133531: contig of 36994 bp in length.  
Location/Qualifiers

FEATURES

source  
1. .170524  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="16"  
/clone="RP23-255N15"  
misc\_feature  
1. .1251  
/note="assembly\_name:Contig14"  
gap  
1252. .1351  
/estimated\_length=unknown  
misc\_feature  
1352. .3283  
/note="assembly\_name:Contig15"  
gap  
3284. .3383  
/estimated\_length=unknown  
misc\_feature  
3384. .4641  
/note="assembly\_name:Contig16"  
gap  
4642. .4741  
/estimated\_length=unknown  
misc\_feature  
4742. .8114  
/note="assembly\_name:Contig18"  
gap  
8115. .8214  
/estimated\_length=unknown  
misc\_feature  
8215. .13181  
/note="assembly\_name:Contig19"  
gap  
13182. .13281  
/estimated\_length=unknown  
misc\_feature  
13282. .16008  
/note="assembly\_name:Contig20"  
gap  
16009. .16108

misc\_feature  
16109. .21115  
/note="assembly\_name:Contig21"  
gap  
21116. .21215  
/estimated\_length=unknown  
misc\_feature  
21216. .24744  
/note="assembly\_name:Contig22"  
gap  
24745. .24844  
/estimated\_length=unknown  
misc\_feature  
24845. .32255  
/note="assembly\_name:Contig24"  
gap  
32256. .32355  
/estimated\_length=unknown  
misc\_feature  
32356. .40226  
/note="assembly\_name:Contig25"  
gap  
40227. .40326  
/estimated\_length=unknown  
misc\_feature  
40327. .47032  
/note="assembly\_name:Contig26"  
gap  
47033. .47132  
/estimated\_length=unknown  
misc\_feature  
47133. .55420  
/note="assembly\_name:Contig27"  
gap  
55421. .55520  
/estimated\_length=unknown  
misc\_feature  
55521. .66858  
/note="assembly\_name:Contig28"  
gap  
66859. .66958  
/estimated\_length=unknown  
misc\_feature  
66959. .79315  
/note="assembly\_name:Contig29  
clone\_end:T7  
vector\_side:left"  
gap  
79316. .79415  
/estimated\_length=unknown  
misc\_feature  
79416. .98103  
/note="assembly\_name:Contig30"  
gap  
98104. .98203  
/estimated\_length=unknown  
misc\_feature  
98204. .133430  
/note="assembly\_name:Contig31  
clone\_end:SP6  
vector\_side:right"  
gap  
133431. .133530  
/estimated\_length=unknown  
misc\_feature  
133531. .170524  
/note="assembly\_name:Contig32"

ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 170524;  
Best Local Similarity 91.3%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAACAATGACCAACTGGGCT 23  
|||||||  
Db 161835 AAAAAACAATGACCACTGGCT 161813

RESULT 30

AC073470  
LOCUS  
DEFINITION Homo sapiens chromosome 7 clone RP11-570M16, WORKING DRAFT  
AC073470 171673 bp DNA linear HTG 04-AUG-2000  
SEQUENCE, 33 unordered pieces.  
ACCESSION AC073470  
VERSION AC073470.3 GI:9690412  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 171673)  
AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 171673)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Aug 4, 2000 this sequence version replaced gi:9309529.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H.NH0570M16  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 151549 bases at least Q40  
Consensus quality: 157491 bases at least Q30  
Consensus quality: 160825 bases at least Q20  
Insert size: 168000; agarose-fp  
Insert size: 168473; sum-of-contigs  
Quality coverage: 3.65 in Q20 bases; agarose-fp  
Quality coverage: 3.73 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1043: contig of 1043 bp in length  
\* 1044: 1143: gap of unknown length  
\* 1144: 2412: contig of 1269 bp in length  
\* 2413: 2512: gap of unknown length  
\* 2513: 3933: contig of 1421 bp in length  
\* 3934: 4033: gap of unknown length  
\* 4034: 5390: contig of 1357 bp in length  
\* 5391: 5491: gap of unknown length  
\* 5491: 6629: contig of 1139 bp in length  
\* 6630: 6729: gap of unknown length  
\* 6730: 7967: contig of 1237 bp in length  
\* 7967: 8066: gap of unknown length  
\* 8067: 9485: contig of 1419 bp in length  
\* 9486: 9585: gap of unknown length  
\* 9586: 10897: contig of 1312 bp in length  
\* 10898: 12498: contig of 1501 bp in length  
\* 12499: 12598: gap of unknown length  
\* 12599: 13735: contig of 1137 bp in length  
\* 13736: 13835: gap of unknown length  
\* 13836: 15277: contig of 1442 bp in length  
\* 15278: 15377: gap of unknown length  
\* 15378: 17363: contig of 1986 bp in length  
\* 17364: 17463: gap of unknown length  
\* 17464: 18722: contig of 1259 bp in length  
\* 18723: 18823: gap of unknown length  
\* 18823: 21225: contig of 2403 bp in length  
\* 21226: 21325: gap of unknown length  
\* 21326: 23191: contig of 1866 bp in length  
\* 23192: 23291: gap of unknown length  
\* 23292: 26263: contig of 2972 bp in length  
\* 26264: 26363: gap of unknown length  
\* 26364: 30191: contig of 3828 bp in length  
\* 30192: 30291: gap of unknown length  
\* 30292: 32484: contig of 2193 bp in length

\* 32485 32584: gap of unknown length  
\* 32585 36044: contig of 3460 bp in length  
\* 36045 36144: gap of unknown length  
\* 36145 41434: contig of 5290 bp in length  
\* 41435 41534: gap of unknown length  
\* 41535 44533: contig of 2999 bp in length  
\* 44534 44633: gap of unknown length  
\* 44634 50067: contig of 5434 bp in length  
\* 50068 50167: gap of unknown length  
\* 50168 55034: contig of 4867 bp in length  
\* 55035 55134: gap of unknown length  
\* 55135 60252: contig of 5118 bp in length  
\* 60253 60352: gap of unknown length  
\* 60353 67994: contig of 7642 bp in length  
\* 67995 68094: gap of unknown length  
\* 68095 74834: contig of 6740 bp in length  
\* 74835 74934: gap of unknown length  
\* 74935 83345: contig of 8411 bp in length  
\* 83346 83445: gap of unknown length  
\* 83446 95023: contig of 11578 bp in length  
\* 95024 95123: gap of unknown length  
\* 95124 103633: contig of 8510 bp in length  
\* 103634 103733: gap of unknown length  
\* 103734 115298: contig of 11565 bp in length  
\* 115299 115398: gap of unknown length  
\* 115399 128887: contig of 13489 bp in length  
\* 128888 128987: gap of unknown length  
\* 128988 145911: contig of 18924 bp in length  
\* 145912 146011: gap of unknown length  
\* 146012 171673: contig of 25662 bp in length.

FEATURES

Location/Qualifiers  
1..171673  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/clone="RP11-570M16"  
1..1043  
/note="assembly\_name:Contig14"  
1044..1143  
/estimated\_length=unknown  
1144..2412  
/note="assembly\_name:Contig28"  
2413..2512  
/estimated\_length=unknown  
2513..3933  
/note="assembly\_name:Contig35"  
3934..4033  
/estimated\_length=unknown  
4034..5390  
/note="assembly\_name:Contig36"  
5391..5490  
/estimated\_length=unknown  
5491..6629  
/note="assembly\_name:Contig37"  
6630..6729  
/estimated\_length=unknown  
6730..7966  
/note="assembly\_name:Contig38"  
7967..8066  
/estimated\_length=unknown  
8067..9485  
/note="assembly\_name:Contig39"  
9486..9585  
/estimated\_length=unknown  
9586..10897  
/note="assembly\_name:Contig40"  
10898..10997  
/estimated\_length=unknown  
10998..12498  
/note="assembly\_name:Contig41"  
12499..12598  
/estimated\_length=unknown  
misc\_feature  
1044..1143  
gap  
1144..2412  
misc\_feature  
2413..2512  
gap  
2513..3933  
misc\_feature  
3934..4033  
gap  
4034..5390  
misc\_feature  
5391..5490  
gap  
5491..6629  
misc\_feature  
6630..6729  
gap  
6730..7966  
misc\_feature  
7967..8066  
gap  
8067..9485  
misc\_feature  
9486..9585  
gap  
9586..10897  
misc\_feature  
10898..10997  
gap  
10998..12498  
misc\_feature  
12499..12598  
gap

misc\_feature 12599..13735  
/notes="assembly\_name:Contig42"  
13736..13835  
/estimated\_length=unknown  
misc\_feature 13836..15277  
/notes="assembly\_name:Contig43"  
15278..15377  
/estimated\_length=unknown  
misc\_feature 15378..17363  
/notes="assembly\_name:Contig44"  
17364..17463  
/estimated\_length=unknown  
misc\_feature 17464..18722  
/notes="assembly\_name:Contig45"  
clone\_end:SP6  
vector\_side:right"  
18723..18822  
/estimated\_length=unknown  
misc\_feature 18823..21225  
/notes="assembly\_name:Contig46"  
21226..21325  
/estimated\_length=unknown  
misc\_feature 21326..23191  
/notes="assembly\_name:Contig47"  
23192..23291  
/estimated\_length=unknown  
misc\_feature 23292..26263  
/notes="assembly\_name:Contig48"  
26264..26363  
/estimated\_length=unknown  
misc\_feature 26364..30191  
/notes="assembly\_name:Contig49"  
30192..30291  
/estimated\_length=unknown  
misc\_feature 30292..32484  
/notes="assembly\_name:Contig50"  
32485..32584  
/estimated\_length=unknown  
misc\_feature 32585..36044  
/notes="assembly\_name:Contig51"  
36045..36144

Query Match 79.2%; Score 19.8; DB 14; Length 171673;  
Best Local Similarity 91.3%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTG 24  
|||||  
Db 115758 AAAAAAATGACCAACAGGGATG 115780  
|||||

RESULT 31  
AC018728/c  
LOCUS AC018728 181935 bp DNA linear PRI 13-JUN-2002  
DEFINITION Homo sapiens chromosome 7 clone RP11-206N24, complete sequence.  
ACCESSION AC018728  
VERSION AC018728.6 GI:21397372  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 181935)  
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.  
and Haugen,E.D.  
Direct Submission  
Unpublished  
REFERENCE 2 (bases 1 to 181935)  
AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1999) Human Genome Center, University of

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 181935)  
Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and  
Olson,M.V.  
Direct Submission  
Submitted (30-AUG-2000) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
4 (bases 1 to 181935)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.  
and Haugen,E.D.  
Direct Submission  
Submitted (13-JUN-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
On Jun 13, 2002 this sequence version replaced gi:9945160.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwghtgs@u.washington.edu  
----- Project Information  
Center project name: chr-7  
Center clone name: RP11-206N24 (djs269)  
----- Summary Statistics  
Assembly program: Phrap; version 0.990319  
Consensus quality: 181883 bases at least Q40  
Consensus quality: 181934 bases at least Q30  
Consensus quality: 181935 bases at least Q20  
Insert size: 181935; sum-of-contigs  
Quality coverage: 10.0x in Q20 bases; sum-of-contigs  
-----

Overlapping Sequences:  
5': RP11-102G17 (UWGC:djs267) AC078842, 133761-bp overlap  
3': RP11-181K7 (UWGC:djs270) AC009517, 4307-bp clone overlap  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an  
alternate chemistry or covered by high quality data (i.e., Phred  
quality >= 30); an attempt was made to resolve all sequencing  
problems, such as compressions and repeats; all regions were  
covered by at least one plasmid subclone or more than one M13  
subclone; and the assembly was confirmed by restriction digest.  
-----  
Sequence Validation:  
This sequence has been validated by Multiple Complete Digest  
fingerprinting. Comparison of the experimentally derived digest  
fragments with sequence-predicted fragments is given below.  
The electronically-digested sequence consists of both insert and  
vector, in order to accurately represent the entire circular BAC.  
Small fragments below a variable cutoff (approximately 400-800 bp)  
are not resolved in the fingerprint and hence do not appear  
in the table. There are no significant remaining discrepancies  
between the experimental and predicted values. Uniquely ordered  
fragments are separated by dashed lines.

Bg11		NsiI		EcoRI	
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt
-----	-----	-----	-----	-----	-----
17153	19443	20363	20416	8696	8905
-----	-----	-----	-----	-----	-----
2067	2099	340	<800	6	<800



5839	5638	481	<800	915	928	2590	2612	23	<800	7123	7037
15389	15229	2871	2817	2190	2318	6365	6393	3713	3799	2315	2229
606	<800	1314	1322	133	<800	714	<800	676	<800	6908	7037
3819	3880	25	<800	223	<800	5839	5956	5582	5360	90	<800
13873	13714	7045	7062	1621	1636	669	<800	796	<800	12239	12115
1176	1163	2024	2044	8577	8641	317	<800	1417	1322	1168	1185
945	959	2748	2817	6552	6504	2638	2612	1330	1322	5227	5246
2616	2612	40	<800	3953	3971	8082	8138	3827	3799	2879	2850
2208	2285	1427	1395	6996	7037	493	<800	7059	7062	689	<800
1422	1381	4849	4848	132	<800	6654	6682	11906	11857	4381	4512
1734	1728	2499	2492	2240	2229	78	<800	5834	5772	2454	2463
1773	1728	7466	7511	4528	4512	2276	2410	2597	2607	266	<800
758	<800	8699	8670	2449	2463	3810	3708	600	<800	4188	4163
4545	4468	2816	2817	6606	6504	81	<800	499	<800	3383	3415
1858	1866	1637	1544	3886	3807	8	<800	2176	2214	4129	4163
2091	2208	585	<800	1378	1356	2192	85	85	<800	175	<800
7187	7164	107	<800	3587	3634	2192	2192	2192	2214	---	---
2138	2285	3758	3799	3099	3103	Query Match 79.2%; Score 19.8; DB 8; Length 181935;					
315	<800	1505	1544	4601	4512	Best Local Similarity 91.3%; Pred. No. 1.9e+02;					
12337	12706	2265	2350	132	<800	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
1362	1381	2796	2817	2636	2652	QY 2 AAAAAAATGACCAACTGGCTG 24					
1222	1228	1173	1139	1669	1636						
8377	8459	2310	2350	4012	4163	Db 175449 AAAAAAATGACCAACAGGATG 175427					
3664	3708	512	<800	6308	6302	AC140950 190084 bp DNA linear PRI 26-SEP-2003					
5399	5445	9486	9423	2024	2047	Pan troglodytes BAC clone RP43-58122 from 7, complete sequence.					
4687	4639	2326	2350	2608	2579	AC140950					
2045	2099	3865	3799	2455	2463	Pan troglodytes					
972	959	3502	3511	47	<800	AC140950.1 GI:28850252					
295	<800	5390	5567	13348	13324	HTG.					
2427	2612	2803	2817	2744	2730	Pan troglodytes (chimpanzee)					
1085	1078	900	904	638	<800	Pan troglodytes					
2591	2612	881	904	1520	1509	Pan troglodytes					
2275	2285	6801	6738	3270	3220	Pan troglodytes					
2414	2410	5345	5360	7440	7465	Pan troglodytes					
3719	3708	1833	1756	5039	4961	Pan troglodytes					
5636	5445	1059	1060	4761	4747	Pan troglodytes					
						Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
						Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;					
						Hominidae; Pan. (bases 1 to 190084)					
						1 (bases 1 to 190084)					
						Walligorski, J., Haakenson, W. and Haglund, K.					
						The sequence of Pan troglodytes BAC clone RP43-58122					
						Unpublished (2001)					
						2 (bases 1 to 190084)					
						Sulston, J.E. and Wilson, R.					
						Sequencing of Pan troglodytes					
						Unpublished (2001)					
						3 (bases 1 to 190084)					
						Waterston, R.					
						Direct Submission					
						Submitted (05-MAR-2003) Department of Genetics, Washington					
						University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA					
						4 (bases 1 to 190084)					
						Wilson, R.					
						Direct Submission					
						Submitted (26-SEP-2003) Department of Genetics, Washington					

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Summary Statistics  
Center project name: C\_PT058122  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (<http://www.resgen.com>) or Pieter de Jong and co-workers at <http://www.bacpac.chori.org>.

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

FEATURES  
source : 1.190084  
          /organism="Pan troglodytes"  
          /mol\_type="genomic DNA"  
          /db\_xref="taxon:9598"  
          /chromosome="7"  
          /map="7"  
          /clone="RP43-58122"  
          /clone\_lib="RPCI-43"  
          106485..106511  
misc\_feature /note="Unresolved homopolymeric repeat."

ORIGIN

Query Match 79.2%; Score 19.8; DB 8; Length 190084;  
Best Local Similarity 91.3%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACAATGACCAACTGGGCTG 24  
|||||  
Db 55534 AAAAACAATGACCAACAGGATG 55556

RESULT 33

CT010460 232469 bp DNA linear HTG 25-JUL-2005  
LOCUS Mus musculus chromosome 12 clone RP23-311G21, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 16 unordered pieces.

CT010460  
ACCESSION CT010460.1 GI:71142028  
VERSION HTG; HTGS PHASE1.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
1 (bases 1 to 232469)  
Sims, S.  
Direct Submission  
Submitted (21-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
----- Project Information  
Center project name: BM311G21  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 226307 bases at least Q40  
Consensus quality: 227941 bases at least Q30  
Consensus quality: 229053 bases at least Q20  
Insert size: 230969; sum-of-contigs  
Insert size: 243194; 4.0% error; agarose-fp  
Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality  
coverage: 4.41x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2300: contig of 2300 bp in length  
\* 2301 2400: gap of 100 bp  
\* 2401 20064: contig of 17664 bp in length  
\* 20065 20164: gap of 100 bp  
\* 20165 32528: contig of 12364 bp in length  
\* 32529 32628: gap of 100 bp  
\* 32629 50388: contig of 17760 bp in length  
\* 50389 50488: gap of 100 bp  
\* 50489 56605: contig of 6117 bp in length  
\* 56606 56705: gap of 100 bp  
\* 56706 72035: contig of 15330 bp in length  
\* 72036 72135: gap of 100 bp  
\* 72136 114525: contig of 42390 bp in length  
\* 114526 114625: gap of 100 bp  
\* 114626 126383: contig of 11758 bp in length  
\* 126384 126483: gap of 100 bp  
\* 126484 142084: contig of 15601 bp in length  
\* 142085 142184: gap of 100 bp  
\* 142185 152571: contig of 10387 bp in length  
\* 152572 152671: gap of 100 bp  
\* 152672 155785: contig of 3114 bp in length  
\* 155786 155885: gap of 100 bp  
\* 155886 170998: contig of 15113 bp in length  
\* 170999 171098: gap of 100 bp  
\* 171099 175427: contig of 4329 bp in length  
\* 175428 175527: gap of 100 bp  
\* 175528 203780: contig of 28253 bp in length  
\* 203781 203880: gap of 100 bp  
\* 203881 206442: contig of 2562 bp in length  
\* 206443 206542: gap of 100 bp  
\* 206543 232469: contig of 25927 bp in length.

FEATURES  
source Location/Qualifiers  
1..232469  
    /organism="Mus musculus"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:10090"  
    /chromosome="12"  
    /clone="RP23-311G21"  
    /clone\_lib="RPCI-23"  
misc\_feature 1..2300

-AUTHORS		TITLE		JOURNAL PUBMED REFERENCE AUTHORS		TITLE JOURNAL		COMMENT	
<div><div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div><div>misc_feature</div></div></div>									

AUTHORS										
JOURNAL										
PUBMED										
REFERENCE										
AUTHORS										

The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Irikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

12869764

2 (bases 1 to 2346)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,

Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,  
Takaku-Akaiira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,  
Yasunishi,A. and Hayashizaki,Y.

FEATURES  
source  
1. .2346  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J023133N22"

## ORIGIN

Query Match 77.6%; Score 19.4; DB 15; Length 2346;  
Best Local Similarity 95.2%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACAATGACCAACTGGGCT 23

DB 1483 AAAATAATGACCAACTGGGCT 1463

## RESULT 35

AP008212\_261/c

## WPCOMMENT

Sequence split into 308 fragments LOCUS AP008212 Accession AP008212

Fragment Name	Begin	End
AP008212_000	1	110000
AP008212_001	100001	210000
AP008212_002	200001	310000
AP008212_003	300001	410000
AP008212_004	400001	510000
AP008212_005	500001	610000
AP008212_006	600001	710000
AP008212_007	700001	810000
AP008212_008	800001	910000
AP008212_009	900001	1010000
AP008212_010	1000001	1110000
AP008212_011	1100001	1210000
AP008212_012	1200001	1310000
AP008212_013	1300001	1410000
AP008212_014	1400001	1510000
AP008212_015	1500001	1610000
AP008212_016	1600001	1710000
AP008212_017	1700001	1810000
AP008212_018	1800001	1910000
AP008212_019	1900001	2010000
AP008212_020	2000001	2110000
AP008212_021	2100001	2210000
AP008212_022	2200001	2310000
AP008212_023	2300001	2410000
AP008212_024	2400001	2510000
AP008212_025	2500001	2610000
AP008212_026	2600001	2710000
AP008212_027	2700001	2810000
AP008212_028	2800001	2910000
AP008212_029	2900001	3010000
AP008212_030	3000001	3110000
AP008212_031	3100001	3210000
AP008212_032	3200001	3310000
AP008212_033	3300001	3410000
AP008212_034	3400001	3510000
AP008212_035	3500001	3610000
AP008212_036	3600001	3710000
AP008212_037	3700001	3810000
AP008212_038	3800001	3910000
AP008212_039	3900001	4010000
AP008212_040	4000001	4110000
AP008212_041	4100001	4210000
AP008212_042	4200001	4310000
AP008212_043	4300001	4410000
AP008212_044	4400001	4510000
AP008212_045	4500001	4610000

AP008212_046	4600001	4710000
AP008212_047	4700001	4810000
AP008212_048	4800001	4910000
AP008212_049	4900001	5010000
AP008212_050	5000001	5110000
AP008212_051	5100001	5210000
AP008212_052	5200001	5310000
AP008212_053	5300001	5410000
AP008212_054	5400001	5510000
AP008212_055	5500001	5610000
AP008212_056	5600001	5710000
AP008212_057	5700001	5810000
AP008212_058	5800001	5910000
AP008212_059	5900001	6010000
AP008212_060	6000001	6110000
AP008212_061	6100001	6210000
AP008212_062	6200001	6310000
AP008212_063	6300001	6410000
AP008212_064	6400001	6510000
AP008212_065	6500001	6610000
AP008212_066	6600001	6710000
AP008212_067	6700001	6810000
AP008212_068	6800001	6910000
AP008212_069	6900001	7010000
AP008212_070	7000001	7110000
AP008212_071	7100001	7210000
AP008212_072	7200001	7310000
AP008212_073	7300001	7410000
AP008212_074	7400001	7510000
AP008212_075	7500001	7610000
AP008212_076	7600001	7710000
AP008212_077	7700001	7810000
AP008212_078	7800001	7910000
AP008212_079	7900001	8010000
AP008212_080	8000001	8110000
AP008212_081	8100001	8210000
AP008212_082	8200001	8310000
AP008212_083	8300001	8410000
AP008212_084	8400001	8510000
AP008212_085	8500001	8610000
AP008212_086	8600001	8710000
AP008212_087	8700001	8810000
AP008212_088	8800001	8910000
AP008212_089	8900001	9010000
AP008212_090	9000001	9110000
AP008212_091	9100001	9210000
AP008212_092	9200001	9310000
AP008212_093	9300001	9410000
AP008212_094	9400001	9510000
AP008212_095	9500001	9610000
AP008212_096	9600001	9710000
AP008212_097	9700001	9810000
AP008212_098	9800001	9910000
AP008212_099	9900001	10010000
AP008212_100	10000001	10110000
AP008212_101	10100001	10210000
AP008212_102	10200001	10310000
AP008212_103	10300001	10410000
AP008212_104	10400001	10510000
AP008212_105	10500001	10610000
AP008212_106	10600001	10710000
AP008212_107	10700001	10810000
AP008212_108	10800001	10910000
AP008212_109	10900001	11010000
AP008212_110	11000001	11110000
AP008212_111	11100001	11210000
AP008212_112	11200001	11310000
AP008212_113	11300001	11410000
AP008212_114	11400001	11510000
AP008212_115	11500001	11610000
AP008212_116	11600001	11710000
AP008212_117	11700001	11810000
AP008212_118	11800001	11910000

11900001 12010000  
12000001 12110000  
12000001 12110000  
12100001 12210000  
12200001 12310000  
12300001 12410000  
12400001 12510000  
12500001 12610000  
12600001 12710000  
12700001 12810000  
12800001 12910000  
12900001 13010000  
13000001 13110000  
13100001 13210000  
13200001 13310000  
13300001 13410000  
13400001 13510000  
13500001 13610000  
13600001 13710000  
13700001 13810000  
13800001 13910000  
13900001 14010000  
14000001 14110000  
14100001 14210000  
14200001 14310000  
14300001 14410000  
14400001 14510000  
14500001 14610000  
14600001 14710000  
14700001 14810000  
14800001 14910000  
14900001 15010000  
15000001 15110000  
15100001 15210000  
15200001 15310000  
15300001 15410000  
15400001 15510000  
15500001 15610000  
15600001 15710000  
15700001 15810000  
15800001 15910000  
15900001 16010000  
16000001 16110000  
16100001 16210000  
16200001 16310000  
16300001 16410000  
16400001 16510000  
16500001 16610000  
16600001 16710000  
16700001 16810000  
16800001 16910000  
16900001 17010000  
17000001 17110000  
17100001 17210000  
17200001 17310000  
17300001 17410000  
17400001 17510000  
17500001 17610000  
17600001 17710000  
17700001 17810000  
17800001 17910000  
17900001 18010000  
18000001 18110000  
18100001 18210000  
18200001 18310000  
18300001 18410000  
18400001 18510000  
18500001 18610000  
18600001 18710000  
18700001 18810000  
18800001 18910000  
18900001 19010000  
19000001 19110000  
19100001 19210000

AP008212\_119  
AP008212\_120  
AP008212\_121  
AP008212\_122  
AP008212\_123  
AP008212\_124  
AP008212\_125  
AP008212\_126  
AP008212\_127  
AP008212\_128  
AP008212\_129  
AP008212\_130  
AP008212\_131  
AP008212\_132  
AP008212\_133  
AP008212\_134  
AP008212\_135  
AP008212\_136  
AP008212\_137  
AP008212\_138  
AP008212\_139  
AP008212\_140  
AP008212\_141  
AP008212\_142  
AP008212\_143  
AP008212\_144  
AP008212\_145  
AP008212\_146  
AP008212\_147  
AP008212\_148  
AP008212\_149  
AP008212\_150  
AP008212\_151  
AP008212\_152  
AP008212\_153  
AP008212\_154  
AP008212\_155  
AP008212\_156  
AP008212\_157  
AP008212\_158  
AP008212\_159  
AP008212\_160  
AP008212\_161  
AP008212\_162  
AP008212\_163  
AP008212\_164  
AP008212\_165  
AP008212\_166  
AP008212\_167  
AP008212\_168  
AP008212\_169  
AP008212\_170  
AP008212\_171  
AP008212\_172  
AP008212\_173  
AP008212\_174  
AP008212\_175  
AP008212\_176  
AP008212\_177  
AP008212\_178  
AP008212\_179  
AP008212\_180  
AP008212\_181  
AP008212\_182  
AP008212\_183  
AP008212\_184  
AP008212\_185  
AP008212\_186  
AP008212\_187  
AP008212\_188  
AP008212\_189  
AP008212\_190  
AP008212\_191

Query Match 77.6%; Score 19.4; DB 15; Length 110000;  
Best Local Similarity 95.2%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAACAATGACCACTGGGCT 23  
|||||  
Db 52673 AAAATAATGACCACTGGGCT 52653  
|||||

RESULT 36  
AP003517/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,  
BAC clone:OSUNBa0062J02.  
ACCESSION  
AP003517  
VERSION  
AP003517.3  
KEYWORDS  
GI:51535415  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE  
1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC  
clone:OSUNBa0062J02  
Published Only in Database (2001)  
2 (bases 1 to 151636)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Submitted (18-APR-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
On Aug 24, 2004 this sequence version replaced gi:46849600.  
Genes were predicted from the integrated results of the following:  
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH  
(http://www.softberry.com/), GeneMark.hmm  
(http://opal.biology.gatech.edu/GeneMark/), Glimmer-M  
(http://www.tigr.org/tdb/glimmer/glmr\_form.html), RiceHMM  
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor  
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4  
(http://globin.cse.psu.edu/html/docs/sim4.html), gap2  
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The  
genomic sequence was searched against NCBI Nonredundant Protein  
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA  
sequence database at RGP or DDBJ. Protein homologues of the coding  
regions were searched against NCBI Nonredundant Protein database  
with BLASTP. ESTs represent the identified cDNA sequences using  
BLASTN with the corresponding DDBJ accession no. and RGP clone ID.  
Full-length cDNAs represent the identified cDNA sequences using  
BLASTN with the corresponding DDBJ accession no.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with full-length cDNA or  
EST homology (covering almost the entire length of partial  
sequence) is classified as an 'unknown' protein. A gene predicted  
by two or more gene prediction programs is classified as a  
'hypothetical' protein according to RGP standard. A gene  
predicted by a single gene prediction program is also classified as  
a probable 'hypothetical' protein and is included as a  
miscellaneous feature of the sequence.  
The orientation of the sequence is from M13rev to -21M13 of the BAC  
clone. This sequence of OSUNBa0062J02 clone has an overlap with  
OJ1226\_A12 (DDBJ: AP004008) clone at 5' end and with P0453H04



Best Local Similarity	95.2%;	Pred. No. 2.9e+02;	Mismatches	20;	Conservative	0;	Indels	1;	Gaps	0;
QY	3	AAAAAATGACCAACTGGGCT 23								
Db	131918	AAAAAATGACCAACTGGGCT 131898								
RESULT 37										
AP005453/c										
LOCUS										
DEFINITION		Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,								
ACCESSION		PAC clone:P0453H04.								
VERSION		AP005453								
KEYWORDS		AP005453.1 GI:21624005								
SOURCE		Oryza sativa (japonica cultivar-group)								
ORGANISM		Oryza sativa (japonica cultivar-group)								
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
AUTHORS		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;								
TITLE		Ehrhartoideae; Oryzaceae; Oryza.								
JOURNAL		1								
REFERENCE		Sasaki, T., Matsumoto, T. and Katayose, Y.								
AUTHORS		Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC								
TITLE		clone:P0453H04								
JOURNAL		Published Only in Database (2002)								
REFERENCE		2 (bases 1 to 175047)								
AUTHORS		Sasaki, T., Matsumoto, T. and Katayose, Y.								
TITLE		Direct Submission								
JOURNAL		Submitted (20-JUN-2002) Takuji Sasaki, National Institute of								
COMMENT		Agrbiological Sciences, Rice Genome Research Program; Kannondai								
		2-1-2, Tsukuba, Ibaraki 305-8602, Japan								
		(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,								
		Tel:81-298-38-7441, Fax:81-298-38-7468)								
		Genes were predicted from the integrated results of the following:								
		GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH								
		(http://www.softberry.com/), GeneMark hmh								
		(http://opal.biology.gatech.edu/GeneMark/), GlimmerM								
		(http://www.tigr.org/tdb/glimmer/glmr_form.html), RiceHMM								
		(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor								
		(http://bioinformatics.laastate.edu/cgi-bin/sp.cgi), sim4								
		(http://globin.cse.psu.edu/html/docs/sim4.html), gap2								
		(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The								
		genomic sequence was searched against NCBI NonRedundant Protein								
		database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA								
		sequence database at RGP or DDBJ. Protein homologs of the coding								
		regions were searched against NCBI NonRedundant Protein database								
		with BLASTP. ESTs represent the identified cDNA sequences using								
		BLASTN with the corresponding DDBJ accession no. and RGP clone ID.								
		Full-length cDNAs represent the identified cDNA sequences using								
		BLASTN with the corresponding DDBJ accession no.								
		A gene with identity or significant homology to a protein is								
		classified based on the protein name to indicate the homology level								
		such as same name, 'putative,' and '-like protein'. A gene without								
		significant homology to any protein but with full-length cDNA or								
		EST homology (covering almost the entire length of partial								
		sequence) is classified as an 'unknown' protein. A gene predicted								
		by two or more gene prediction programs is classified as a								
		'hypothetical' protein according to IRGSP standard. A gene								
		predicted by a single gene prediction program is also classified as a								
		probable 'hypothetical' protein and is included as a								
		miscellaneous feature of the sequence.								
		The orientation of the sequence is from SP6 to T7 of the PAC clone.								
		This sequence of P0453H04 clone has an overlap with OSUNBA0062J02								
		(DDBJ: AP003517) clone at 5' end and with OSUNBA0085J13 (DDBJ:								
		AP003565) at 3' end. Detailed information on overlap and assembly								
		quality together with annotation of this entry is available at								
		http://rgp.dna.affrc.go.jp/GenomeSeq.html.								
FEATURES		Location/Qualifiers								
source		1. .175047								
		/organism="Oryza sativa (japonica cultivar-group)"								
		/mol_type="genomic DNA"								
		/cultivar="Nipponbare"								

gene	/db_xref="taxon:39947"	
	/chromosome="6"	
	/clone="P0453H04"	
	complement(2482..2691)	
	/gene="P0453H04.1"	
	complement(2482..2691)	
	/gene="P0453H04.1"	
	/note="hypothetical ORF	
	predicted by glimmerM	
	this category is not included in IRGSP standard"	
	complement(join(3217..3678,5673..5759,5979..6023))	
	/gene="P0453H04.2"	
	complement(join(3217..3678,5673..5759,5979..6023))	
	/gene="P0453H04.2"	
	/note="hypothetical ORF	
	predicted by GeneMark.hmm	
	this category is not included in IRGSP standard"	
	join(7151..7301,7947..8340,8427..8556)	
	/gene="P0453H04.3"	
	join(7151..7301,7947..8340,8427..8556)	
	/gene="P0453H04.3"	
	/note="hypothetical ORF	
	predicted by GENSCAN	
	this category is not included in IRGSP standard"	
	complement(8579..9350)	
	/gene="P0453H04.4"	
	complement(join(<8579..8890,9342..>9350))	
	/gene="P0453H04.4"	
	/note="start and end point are not identified"	
	complement(join(8579..8890,9342..9350))	
	/gene="P0453H04.4"	
	/note="predicted by GENSCAN etc."	
	/codon_start=1	
	/product="hypothetical protein"	
	/protein_id="BAD38098.1"	
	/db_xref="GI:51534974"	
	/translation="MTTKDEQCRPAGGSRSDGDRWAEGWRAAGRGSGRGAAGGGE	
	MTGGRRRDENAAAREEASSSRPVEEAARVGREGGCAAVRGGRQPMGKRKSGGSP	
	TSPR"	
	complement(9530..9809)	
	/gene="P0453H04.5"	
	complement(join(<9530..9605,9700..>9809))	
	/gene="P0453H04.5"	
	/note="start and end point are not identified"	
	complement(join(9530..9605,9700..9809))	
	/gene="P0453H04.5"	
	/note="Predicted by GeneMark.hmm etc."	
	/codon_start=1	
	/product="hypothetical protein"	
	/protein_id="BAD38099.1"	
	/db_xref="GI:51534975"	
	/translation="WTVLCSMFLSVAAPFDGSFEAAAKVLAFVAGGDLMGAVAVASC	
	HGADPSPGQAGRPRA"	
	10192..18908	
	/gene="P0453H04.6"	
	join(10192..10419,11280..11361,11889..11967,12067..12164,	
	12292..12348,12932..13019,13129..13238,13421..13522,	
	13777..13928,14729..14839,14915..14984,15075..15114,	
	15528..15692,15794..15870,17190..17306,17445..17710,	
	18230..18325,18409..18475,18568..18908)	
	/gene="P0453H04.6"	
	/note="supported by full-length cDNA(s) : AK072591"	
	join(10286..10419,11280..11361,11889..11967,12067..12164,	
	12292..12348,12932..13019,13129..13238,13421..13522,	
	13777..13928,14729..14839,14915..14984,15075..15114,	
	15528..15692,15794..15870,17190..17306,17445..17710,	
	18230..18325,18409..18475,18568..18621)	
	/gene="P0453H04.6"	
	/note="ATA binding protein associated factor 5	
	contains full-length cDNA(s) : AK072591"	
	/codon_start=1	
	/product="putative TAF5"	
	/protein_id="BAD38100.1"	



```
/db_xref="GI:51534976"
/translation="MEDEMEKKVQOYLQKGRFLTELALQERNRISTSSVDVALA
RENDPARYDGYSLRTWAYSDQYKHELLRLVLPVFIHSFMDLVAEGHTQEARSE
FTFHEHMSRLQKLEGLSPSHSEMLARSLRNQRFRIKCEYSYELLQYL
QKQALVDVIGIINERTFFVSPQPSLISDDTDVVALVGTKKDLAKQINLKEVHWGLL
EDSVEMKTLLESDEKTEAESKDAEDNNRKSSEGQGGVKKVKKDKIAGATG
KTKNSTSVSHAPRVKPELTLPVPEVEQSTLEDLRNRAOLNSLALPSVSVFTFLN
TINGLNCSSISHDGSLVGGFSDSVKVMDSKIGOPPKTSSPOGENGLSQGVERTSAS
DVGKRPYTLFQHGSGPVSAARSPFGDFLLSSSDSTIRLWSTKLNANLVCKGNYP
VMDVQSPVGHYFASASHDKTAKIMSKIQPIRMWAGHSVLSVDCVQHWVNCYIATG
SDKTVRLVDVQTEGICIRFIHRSMVLKSLAMSPDGRYMASGDEGDTIMMDLSSGRC
VSPLGSSCVMSLAYSCGALIASGADCTVKLMDVASSTKVLKTDSTTNRLRLMK
TLTKTGPVYTLRFSRNLFAAGALSIGS"
complement (20251..20484)
/gene="P0453H04.7"
complement (<20251..>20484)
/genes="P0453H04.7"
/notes="start and end point are not identified"
complement (20251..20484)
/gene="P0453H04.7"
/notes="predicted by GlimmerM etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD38101.1"
/db_xref="GI:51534977"
/translation="MFGHVYVAGISIQAGTGGAVRQGRCAISPGSGQAGVDASG
ARQGGQAPCDRGLGKCGWFSATGLGLVVAGH"
complement (join(21526..21633,22311..22403,25381..25499,
28308..28323,28773..28832))
/genes="P0453H04.8"
complement (join(21526..21633,22311..22403,25381..25499,
28308..28323,28773..28832))
/gene="P0453H04.8"
/notes="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
32254..32421
/gene="P0453H04.9"
32254..32421
/gene="P0453H04.9"
/notes="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
35423..41646
/gene="P0453H04.10"
join(<35423..36233,36603..36722,37003..37116,41025..41143,
41257..>41646)
/gene="P0453H04.10"
/notes="supported by full-length cDNA(s): AK103769"
35423..41646
/gene="P0453H04.10"
/notes="contains full-length cDNA(s): AK103769
non-coding transcript
probably inactive due to including stop codon(s) in CDS"
complement (join(41799..41902,44014..44110))
/gene="P0453H04.11"
complement (join(41799..41902,44014..44110))
/gene="P0453H04.11"
/notes="hypothetical ORF
predicted by RiceHMM
this category is not included in IRGSP standard"
44222..46867
/gene="P0453H04.12"
<44222..>46867

Query Match 77.6%; Score 19.4; DB 15; Length 175047;
Best Local Similarity 95.2%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 AAAACAATGACCAACTGGGCT 23
|||||
17213 AAAATAATGACCAACTGGGCT 17193

Qy Db
```

```
RESULT 38
CQ337709
LOCUS CQ337709 600 bp DNA linear PAT 23-JAN-2004
DEFINITION Sequence 11803 from Patent WO0157275.
ACCESSION CQ337709
VERSION CQ337709.1 GI:41286780
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1
AUTHORS Penn,S.G., Hanzel,D.K., Chen,W. and Rank,D.R.
TITLE Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human brain
JOURNAL Patent: WO 0157275-A 11803 09-AUG-2001;
Aeomica, Inc. (US)
FEATURES
Location/Qualifiers
1..600
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="MAP TO AC009178.4
EXPRESSED IN BRAIN, SIGNAL = 1.3"
ORIGIN
Query Match 76.8%; Score 19.2; DB 6; Length 600;
Best Local Similarity 87.5%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy Db 370 AAAAACCATGACCAACTGGTTGT 393
|||||
2 AAAACAATGACCAACTGGGCTGT 25
|||||

RESULT 39
AK102491
LOCUS AK102491 1310 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J033094019, full
insert sequence.
ACCESSION AK102491
VERSION AK102491.1 GI:32987700
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-length cDNA Project Team,
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiiki,T., Ohneda,E., Yabagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Otono,Y., Teunoda,Y.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narioka,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
Tkeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsuura,K., RIKEN,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,K.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
PUBMED 12869764
REFERENCE 2 (bases 1 to 1310)
```

## AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imoto,A., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kuroseki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Mateubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Nikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

## TITLE

Direct Submission  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@ias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cdna/>

NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kuroseki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Nikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center

and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

Location/Qualifiers

1. .1310

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="J033094019"

## FEATURES

source

## ORIGIN

source

Query Match 76.8%; Score 19.2; DB 15; Length 1310;  
Best Local Similarity 87.5%; Pred. No. 4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTGT 25

|||||

DB 1229 AAAAAAGTGACCAACTGGGCTGT 1252

|||||

## RESULT 40

AP004334

LOCUS

DEFINITION

AP004334

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

AP004334 101607 bp DNA linear  
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,  
PAC clone:P0455H11.

AP004334 GI:21203158

Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

1 Sasaki,T., Matsumoto,T. and Yamamoto,K.

Oryza sativa nipponbare (G3) genomic DNA, chromosome 7, PAC

clone:P0455H11

Published Only in Database (2001)

2 (bases 1 to 101607)

Sasaki,T., Matsumoto,T. and Yamamoto,K.

Direct Submission

Submitted (14-NOV-2001) Takuji Sasaki, National Institute of

Agrobiological Sciences, Rice Genome Research Program; Kannondai

2-1-2, Tsukuba, Ibaraki, 305-8602, Japan

(E-mail:tsasaki@ias.affrc.go.jp, URL:<http://rgp.dna.affrc.go.jp/>,  
Tel:81-298-38-7441, Fax:81-298-38-7468)

On May 24, 2002 this sequence version replaced gi:16930119.

Genes were predicted from the integrated results of the following:

GENSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), FGENSEH

(<http://www.softberry.com/>), GeneMark.hmm

(<http://opal.biology.gatech.edu/GeneMark/>), GlimmerM

(<http://opal.biology.gatech.edu/glimmer/glmr form.html>), RiceHMM

(<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor

(<http://bioinformatics.iastate.edu/cgi-bin/sp.cgi>), sim4

(<http://globin.cse.psu.edu/html/docs/sim4.html>), gap2

(<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The

genomic sequence was searched against NCBI NonRedundant Protein

database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA

sequence database at RGP or DDBJ. Protein homologues of the coding

regions were searched against NCBI NonRedundant Protein database

with BLASTP. ESTs represent the identified cDNA sequences using

BLASTN with the corresponding DDBJ accession no. and RGP clone ID.

Full-length cDNAs represent the identified cDNA sequences using

BLASTN with the corresponding DDBJ accession no.

A gene with identity or significant homology to a protein is

classified based on the protein name to indicate the homology level

such as same name, 'putative', and '-like protein'. A gene without

significant homology to any protein but with full-length cDNA or

EST homology (covering almost the entire length of partial

sequence) is classified as an 'unknown' protein. A gene predicted

by two or more gene prediction programs is classified as a

'hypothetical' protein according to IRGSP standard. A gene

predicted by a single gene prediction program is also classified as

a probable 'hypothetical' protein and is included as a

miscellaneous feature of the sequence.

The orientation of the sequence is from T7 to SP6 of the PAC clone.

This sequence of P0455H11 clone has an overlap with P0519E12

clone (DDBJ: AP004339) at 5' end and an overlap with OJ1332\_C12

clone (DDBJ: AP003752) at 3' end. Detailed information on overlap

and assembly quality together with annotation of this entry is

available at

<http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

Location/Qualifiers

1. .101607

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="7"

/clone="P0455H11"

Join(4486..4493,5998..6098,7842..7909)

/gene="P0455H11.101"

Join(4486..4493,5998..6098,7842..7909)

misc\_feature





gap /estimated\_length=unknown  
20776..20875  
gap /estimated\_length=unknown  
33308..33407  
gap /estimated\_length=unknown  
43459..43558  
gap /estimated\_length=unknown  
58213..58312  
gap /estimated\_length=unknown  
73363..73462  
gap /estimated\_length=unknown  
85518..85617  
gap /estimated\_length=unknown  
98209..98308  
gap /estimated\_length=unknown  
115205..115304  
gap /estimated\_length=unknown  
131266..131365  
gap /estimated\_length=unknown  
160956..161055  
gap /estimated\_length=unknown  
202836..202935  
gap /estimated\_length=unknown  
247059..247158  
gap /estimated\_length=unknown  
292711..292810  
gap /estimated\_length=unknown

## ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTG 24  
||||||| |

Db 98079 AAAAAACAACAGCAACTGGGCTG 98102

RESULT 42  
AP008213\_262

WPCOMMENT

Sequence split into 297 fragments LOCUS AP008213 Accession AP008213

Fragment Name	Begin	End
AP008213_000	1	110000
AP008213_001	100001	210000
AP008213_002	200001	310000
AP008213_003	300001	410000
AP008213_004	400001	510000
AP008213_005	500001	610000
AP008213_006	600001	710000
AP008213_007	700001	810000
AP008213_008	800001	910000
AP008213_009	900001	1010000
AP008213_010	1000001	1110000
AP008213_011	1100001	1210000
AP008213_012	1200001	1310000
AP008213_013	1300001	1410000
AP008213_014	1400001	1510000
AP008213_015	1500001	1610000
AP008213_016	1600001	1710000
AP008213_017	1700001	1810000
AP008213_018	1800001	1910000
AP008213_019	1900001	2010000
AP008213_020	2000001	2110000
AP008213_021	2100001	2210000
AP008213_022	2200001	2310000
AP008213_023	2300001	2410000
AP008213_024	2400001	2510000
AP008213_025	2500001	2610000
AP008213_026	2600001	2710000
AP008213_027	2700001	2810000
AP008213_028	2800001	2910000
AP008213_029	2900001	3010000

3000001  
3100001  
3210000  
3300001  
3400001  
3500001  
3600001  
3700001  
3800001  
3900001  
4000001  
4100001  
4200001  
4300001  
4400001  
4500001  
4600001  
4700001  
4800001  
4900001  
5000001  
5100001  
5200001  
5300001  
5400001  
5500001  
5600001  
5700001  
5800001  
5900001  
6000001  
6100001  
6200001  
6300001  
6400001  
6500001  
6600001  
6700001  
6800001  
6900001  
7000001  
7100001  
7200001  
7300001  
7400001  
7500001  
7600001  
7700001  
7800001  
7900001  
8000001  
8100001  
8200001  
8300001  
8400001  
8500001  
8600001  
8700001  
8800001  
8900001  
9000001  
9100001  
9200001  
9300001  
9400001  
9500001  
9600001  
9700001  
9800001  
9900001  
10000001  
10100001  
10200001  
10300001

AP008213\_030  
AP008213\_031  
AP008213\_032  
AP008213\_033  
AP008213\_034  
AP008213\_035  
AP008213\_036  
AP008213\_037  
AP008213\_038  
AP008213\_039  
AP008213\_040  
AP008213\_041  
AP008213\_042  
AP008213\_043  
AP008213\_044  
AP008213\_045  
AP008213\_046  
AP008213\_047  
AP008213\_048  
AP008213\_049  
AP008213\_050  
AP008213\_051  
AP008213\_052  
AP008213\_053  
AP008213\_054  
AP008213\_055  
AP008213\_056  
AP008213\_057  
AP008213\_058  
AP008213\_059  
AP008213\_060  
AP008213\_061  
AP008213\_062  
AP008213\_063  
AP008213\_064  
AP008213\_065  
AP008213\_066  
AP008213\_067  
AP008213\_068  
AP008213\_069  
AP008213\_070  
AP008213\_071  
AP008213\_072  
AP008213\_073  
AP008213\_074  
AP008213\_075  
AP008213\_076  
AP008213\_077  
AP008213\_078  
AP008213\_079  
AP008213\_080  
AP008213\_081  
AP008213\_082  
AP008213\_083  
AP008213\_084  
AP008213\_085  
AP008213\_086  
AP008213\_087  
AP008213\_088  
AP008213\_089  
AP008213\_090  
AP008213\_091  
AP008213\_092  
AP008213\_093  
AP008213\_094  
AP008213\_095  
AP008213\_096  
AP008213\_097  
AP008213\_098  
AP008213\_099  
AP008213\_100  
AP008213\_101  
AP008213\_102

AP008213\_103 10410000  
 AP008213\_104 10510000  
 AP008213\_105 10500001  
 AP008213\_106 10710000  
 AP008213\_107 10700001  
 AP008213\_108 10800001  
 AP008213\_109 10900001  
 AP008213\_110 11000001  
 AP008213\_111 11100001  
 AP008213\_112 11200001  
 AP008213\_113 11300001  
 AP008213\_114 11400001  
 AP008213\_115 11500001  
 AP008213\_116 11600001  
 AP008213\_117 11700001  
 AP008213\_118 11800001  
 AP008213\_119 11900001  
 AP008213\_120 12000001  
 AP008213\_121 12100001  
 AP008213\_122 12200001  
 AP008213\_123 12300001  
 AP008213\_124 12400001  
 AP008213\_125 12500001  
 AP008213\_126 12600001  
 AP008213\_127 12700001  
 AP008213\_128 12800001  
 AP008213\_129 12900001  
 AP008213\_130 13000001  
 AP008213\_131 13100001  
 AP008213\_132 13200001  
 AP008213\_133 13300001  
 AP008213\_134 13400001  
 AP008213\_135 13500001  
 AP008213\_136 13600001  
 AP008213\_137 13700001  
 AP008213\_138 13800001  
 AP008213\_139 13900001  
 AP008213\_140 14000001  
 AP008213\_141 14100001  
 AP008213\_142 14200001  
 AP008213\_143 14300001  
 AP008213\_144 14400001  
 AP008213\_145 14500001  
 AP008213\_146 14600001  
 AP008213\_147 14700001  
 AP008213\_148 14800001  
 AP008213\_149 14900001  
 AP008213\_150 15000001  
 AP008213\_151 15100001  
 AP008213\_152 15200001  
 AP008213\_153 15300001  
 AP008213\_154 15400001  
 AP008213\_155 15500001  
 AP008213\_156 15600001  
 AP008213\_157 15700001  
 AP008213\_158 15800001  
 AP008213\_159 15900001  
 AP008213\_160 16000001  
 AP008213\_161 16100001  
 AP008213\_162 16200001  
 AP008213\_163 16300001  
 AP008213\_164 16400001  
 AP008213\_165 16500001  
 AP008213\_166 16600001  
 AP008213\_167 16700001  
 AP008213\_168 16800001  
 AP008213\_169 16900001  
 AP008213\_170 17000001  
 AP008213\_171 17100001  
 AP008213\_172 17200001  
 AP008213\_173 17300001  
 AP008213\_174 17400001  
 AP008213\_175 17500001

AP008213\_176 17600001  
 AP008213\_177 17700001  
 AP008213\_178 17800001  
 AP008213\_179 17900001  
 AP008213\_180 18000001  
 AP008213\_181 18100001  
 AP008213\_182 18200001  
 AP008213\_183 18300001  
 AP008213\_184 18400001  
 AP008213\_185 18500001  
 AP008213\_186 18600001  
 AP008213\_187 18700001  
 AP008213\_188 18800001  
 AP008213\_189 18900001  
 AP008213\_190 19000001  
 AP008213\_191 19100001  
 AP008213\_192 19200001  
 AP008213\_193 19300001  
 AP008213\_194 19400001  
 AP008213\_195 19500001  
 AP008213\_196 19600001  
 AP008213\_197 19700001

Query Match 76.8%; Score 19.2; DB 15; Length 110000;  
 Best Local Similarity 87.5%; Pred. NO. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25  
 DB 70459 AAAAAGTGTGACCAACTGGGCTGT 70482

RESULT 43

AC152963

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC152963 111804 bp DNA linear HTG 03-DEC-2004  
 Takifugu rubripes clone MRC-195B17, WORKING DRAFT SEQUENCE.  
 AC152963  
 HTG: HTGS\_PHASE2; HTGS\_DRAFT.  
 Takifugu rubripes (Fugu rubripes)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontoidea; Tetraodontidae; Takifugu.  
 1 (bases 1 to 111804)  
 Cheng, J.-F., Hamilton, M., Peng, Y., Hosseini, R., Peng, Z., Malinov, I.,  
 and Rubin, E.M.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 111804)  
 Cheng, J.-F., Hamilton, M., Peng, Y., Hosseini, R., Peng, Z., Malinov, I.,  
 and Rubin, E.M.  
 Direct Submission  
 Submitted (25-NOV-2004) Genome Sciences, Lawrence Berkeley National  
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA  
 3 (bases 1 to 111804)  
 Cheng, J.-F., Hamilton, M., Peng, Y., Hosseini, R., Peng, Z., Malinov, I.,  
 and Rubin, E.M.  
 Direct Submission  
 Submitted (03-DEC-2004) Genome Sciences, Lawrence Berkeley National  
 Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

Sequence Produced by Berkeley PGA  
 Web site: <http://pga.lbl.gov>  
 Center Code: PGABERK  
 Center Project Name: F072  
 Bac Clone Name: MRC-195B17

This sequence has been compared to sequences of other species  
 using Vista (<http://www-gsd.lbl.gov/VISTA>). The results can be  
 viewed at:  
[http://pga.lbl.gov/cgi-bin/search\\_cvcgd?type=n&value=PTK7](http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=PTK7)



The order-orientation of the draft sequence was accomplished by using:  
Avid (<http://baboon.math.berkeley.edu/mavid>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

#### Summary Statistics:

Sequencing vector: Plasmid; pUC18

Chemistry: Dye-terminator Big Dye

Assembly program: Phrap version 0.990329.

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 111804: contig of 111804 bp in length.

\* Location/Qualifiers

#### FEATURES

source

1. 111804

/organism="Takifugu rubripes"

/mol\_type="genomic DNA"

/db\_xref="taxon:31033"

/clone="MRC-195B17"

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 111804;  
Best Local Similarity 87.5%; Pred. NO. 3.7e-02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACAATGACCACTGGCGTG 24

Db 81336 AAAGAACAATGACCACTGGCGTG 81359

#### RESULT 44

AC117517/c

LOCUS

DEFINITION Homo sapiens X BAC RP13-356M24 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

AC117517

AC117517.7 GI:21263143

HTG.

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 121628)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Albrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brivet,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Loulesged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogwu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sleson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vaquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

#### TITLE

Direct Submission

Unpublished

2 (bases 1 to 121628)

Worley,K.C.

Direct Submission

Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 121628)

Worley,K.C.

Direct Submission

Submitted (26-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 121628)

Worley,K.C.

Direct Submission

Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 121628)

Worley,K.C.

Direct Submission

Submitted (24-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 121628)

Worley,K.C.

Direct Submission

Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 30, 2002 this sequence version replaced gi:21217391.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

#### COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STGs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.



SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

```

FEATURES
  source      Location/Qualifiers
              1..121628
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="X"
                /clone="RP13-356M24"
                complement(1..2048)
                /notes="overlaps bases 2..2049 of clone AC004673"
                /functions="clone overlap"
                complement(10..291)
                /rpt_family="L1"
                283..361
                /rpt_family="L1MA10"
                362..395
                /rpt_family="CA)n"
                396..548
                /rpt_family="L1MA10"
                744..1126
                /rpt_family="MSTD"
                complement(1573..1792)
                /rpt_family="L2"
                complement(2105..2165)
                /rpt_family="MERSA"
                2547..2786
                /rpt_family="LTR16C"
                2993..3062
                /rpt_family="L2"
                3080..3512
                /rpt_family="L2"
                complement(3513..3801)
                /rpt_family="AluY"
                3802..4211
                /rpt_family="L2"
                complement(4209..4323)
                /rpt_family="L2"
                6708..6819
                /standard_name="47416"
                6759..6925
                /standard_name="42458"
                6763..6872
                /standard_name="84773"
                complement(6993..7308)
                /rpt_family="AluYb8"
                complement(8308..8569)
                /rpt_family="L2"
                8704..8739
                /rpt_family="T-rich"
                complement(8740..9051)
                /rpt_family="AluX"
                complement(9062..10864)
                /rpt_family="L1PA16"
                10870..10986
                /rpt_family="AluDb"
                10987..11017
                /rpt_family="(TAAA)n"
                complement(11028..12468)

```

```

repeat_region  /rpt_family="L1PA16"
complement(12469..12777)
repeat_region  /rpt_family="AluY"
12778..12905
repeat_region  /rpt_family="MER75"
complement(12922..13566)
repeat_region  /rpt_family="L1PA16"
13567..14142
repeat_region  /rpt_family="MLT2D"
complement(14143..14532)
repeat_region  /rpt_family="L1PA16"
14533..15022
repeat_region  /rpt_family="MER75"
complement(15051..15434)
repeat_region  /rpt_family="L1PA16"
complement(15395..15772)
repeat_region  /rpt_family="L1PA13"
complement(15779..16579)
repeat_region  /rpt_family="L1PA13"
16770..16969
repeat_region  /rpt_family="MIR"

```

Query Match 76.8%; Score 19.2; DB 8; Length 121628;  
 Best Local Similarity 87.5%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 2 AAAACAATGACCAACTGGCTGT 25
Db 3517 AAAACAATGATAAACTGGACTGT 3494

```

## RESULT 45

```

AC144797      129045 bp  DNA  linear  ROD 15-MAY-2004
LOCUS      Mus musculus BAC clone RP24-349B2 from chromosome 16, complete
DEFINITION      sequence.
ACCESSION      AC144797
VERSION      AC144797.2 GI:45237305
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 129045)
AUTHORS      VanBrunt,A., Van Brunt,A., Bielicki,L., Meyer,R. and Haakenson,W.
TITLE      The sequence of Mus musculus BAC clone RP24-349B2
JOURNAL      Unpublished (2001)
REFERENCE      2 (bases 1 to 129045)
AUTHORS      Wilson,R.K.
TITLE      Direct Submission
JOURNAL      Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE      3 (bases 1 to 129045)
AUTHORS      Wilson,R.K.
TITLE      Direct Submission
JOURNAL      Submitted (06-MAR-2004) Genome Sequencing Center, 4444 Forest Park
REFERENCE      4 (bases 1 to 129045)
AUTHORS      Wilson,R.K.
TITLE      Direct Submission
JOURNAL      Submitted (15-MAY-2004) Genome Sequencing Center, 4444 Forest Park
COMMENT      On Mar 6, 2004 this sequence version replaced gi:30911150.
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
Summary Statistics
Center project name: M_BB0349B02

```

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

## FEATURES

```

source
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="16"
    /map="16"
    /clone="RP24-349B2"
    /clone_lib="RPCI-24"
    /size="28309"
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=33.19 / Sec struct
    43749. .43816
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=18.82 / Sec struct
    Sc=1.22)"
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=33.19 / Sec struct
    Sc=10.52)"
    /notes="CpG island (%GC=75.0, o/e=0.98, #CpGs=206)"
    /size="103525"
    /notes="Sequence derived from PCR product of project DNA."
    /size="103791"
    /notes="Sequence derived from PCR product of project DNA."
    /size="105003"
    /notes="Sequence derived from one plasmid subclone."
    /size="107451"
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=32.43 / Sec struct
    Sc=7.29)"
    /size="113983"
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=40.24 / Sec struct
    Sc=7.11)"
    /complement(118864..118936)
    /product="tRNA-Ser"
    /notes="Likely pseudogene (HMM Sc=39.14 / Sec struct
    Sc=-14.97)"

```

## ORIGIN

```

Query Match          76.8%; Score 19.2; DB 9; Length 129045;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAACAATGACCAACTGGGCTG 24
    ||||| ||||| ||||| |||||
Db 42068 AAAAAAAGTCCAACTGGGCTG 42091

RESULT 46
AC135550 129968 bp DNA linear HTG 25-JAN-2003
LOCUS Pan troglodytes clone RP43-65H8, WORKING DRAFT SEQUENCE, 4 ordered
DEFINITION pieces.
AC135550 129968 bp DNA linear HTG 25-JAN-2003
VERSION AC135550.2 GI:27901840
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Pan.
REFERENCE 1 (bases 1 to 129968)
AUTHORS Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Hin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., McDowell,J.,
Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE 2 (bases 1 to 129968)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 129968)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT On Jan 25, 2003 this sequence version replaced gi:24110979.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc\_zoo@nhgri.nih.gov
----- Project Information
Center project name: cmd
Center clone name: 065H08

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

```

----- Summary Statistics
Sequencing vector: plasmid, n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 129301 bases at least Q40
Consensus quality: 129516 bases at least Q30
Consensus quality: 129643 bases at least Q20
Insert size: 129000; agarose-fp
Insert size: 129668; sum-of-contigs
Quality coverage: 10.58x in Q20 bases; agarose-fp

```

Quality coverage: 10.52x in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 6330: contig of 6330 bp in length  
 \* 6331 6430: gap of unknown length  
 \* 6431 9836: contig of 3406 bp in length  
 \* 9837 9936: gap of unknown length  
 \* 9937 122751: contig of 112815 bp in length  
 \* 122752 122851: gap of unknown length  
 \* 122852 122968: contig of 7117 bp in length.

## FEATURES

source  
 1..122968  
 /organism="pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="RP43-65H8"  
 /clone\_lib="RP43"

## misc\_feature

1..6330  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 6331..6430

## gap

/estimated\_length=unknown

## misc\_feature

6431..9836  
 /note="assembly\_fragment"  
 9837..9936

## gap

/estimated\_length=unknown

## misc\_feature

9937..122751  
 /note="assembly\_fragment"  
 79726..122968

## misc\_feature

/note="clone overlaps with GenBank Accession Number  
 AC117936 clone RP43-90G9 (center project name cmc)"  
 122752..122851

## gap

/estimated\_length=unknown

## misc\_feature

122852..122968  
 /note="assembly\_fragment  
 clone\_end:17  
 vector\_side:right"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 122968;  
 Best Local Similarity 87.5%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCACTGGGCTG 24  
 ||||| ||||| ||||| ||||| |||||

Db 36236 AAAAAACATGACCTACTGGGCTG 36259  
 ||||| ||||| ||||| ||||| |||||

## RESULT 47

AC144765/c 130836 bp DNA linear PLN 14-JUN-2005  
 LOCUS Medicago truncatula clone mth2-35011, complete sequence.  
 DEFINITION AC144765

ACCESSION AC144765

VERSION AC144765.19 GI:67625917

KEYWORDS HTG.

SOURCE Medicago truncatula (barrel medic)

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 130836)

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 21; Conservative

## 0; Mismatches

## 3; Indels

## 0; Gaps

## 0;

## QY

## 1

## AAAAAACATGACCACTGGGCTG

## 24

## Db

## 70159

## AAAAAACATGACCAAGTGGCTG

## 70136

## RESULT 48

## AC145828

## LOCUS

## DEFINITION

## AC145828

## Pan troglodytes

## BAC clone

## RP43-7E4

## from 7, complete sequence.

## AC145828

## AC145828

Medicago truncatula BAC Clone mth2-35011

Unpublished

2 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (16-MAY-2003) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

3 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (06-MAY-2005) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

4 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (07-MAY-2005) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

5 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (10-MAY-2005) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

6 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (14-MAY-2005) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

7 (bases 1 to 130836)

Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

Cook,D., Kim,D. and Roe,B.A.

Direct Submission

Submitted (14-JUN-2005) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

On Jun 14, 2005 this sequence version replaced gi:65133652.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

----- Location/Qualifiers

1..130836

/organism="Medicago truncatula"

/mol\_type="genomic DNA"

/db\_xref="taxon:3880"

/clone="mth2-35011"

/clone\_lib="Medicago truncatula BAC library H2"

Query Match 76.8%; Score 19.2; DB 15; Length 130836;

Best Local Similarity 87.5%; Pred. No. 3.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCACTGGGCTG 24

||||| ||||| ||||| ||||| |||||

Db 70159 AAAAAACATGACCAAGTGGCTG 70136

||||| ||||| ||||| ||||| |||||

RESULT 48

AC145828

LOCUS

DEFINITION

AC145828

139712 bp DNA linear PRI 26-SEP-2003

Pan troglodytes BAC clone RP43-7E4 from 7, complete sequence.

AC145828

```
AC145828.1 GI:33342460
HTG.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
1 (bases 1 to 139712)
Hodges, J., Kozlowski, A. and Haakensen, W.
The sequence of Pan troglodytes BAC clone RP43-7E4
Unpublished (2001)
2 (bases 1 to 139712)
Sulston, J.E. and Wilson, R.
Sequencing of Pan troglodytes
Unpublished (2001)
3 (bases 1 to 139712)
Wilson, R.K.
Direct Submission
Submitted (30-JUL-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 139712)
Wilson, R.
Direct Submission
Submitted (26-SEP-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: C_PT007E04
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, Clint, Yerkes #C0471; birthdate: 6-6-80). The
clone and detailed information can be obtained from Resgen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. .139712
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clone="RP43-7E4"
/misc_feature 122939..123055

AC092338
Homo sapiens chromosome 16 clone RP11-21M24, complete sequence.
AC092338
Homo sapiens (human)
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 13, 2003 this sequence version replaced gi:18481997.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .143126
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-21M24"

AC092338
Homo sapiens chromosome 16 clone RP11-21M24, complete sequence.
AC092338
Homo sapiens (human)
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 13, 2003 this sequence version replaced gi:18481997.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .143126
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-21M24"

AAACAAATGACCAACTGGGCTG 25
|||||
57835 AAACCAATGACCAACTGGTTGT 57858
|||||

RESULT 49
AC092338
LOCUS
DEFINITION
AC092338
Homo sapiens chromosome 16 clone RP11-21M24, complete sequence.
AC092338
Homo sapiens (human)
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 143126)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 143126)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 13, 2003 this sequence version replaced gi:18481997.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. .143126
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-21M24"

AAACAAATGACCAACTGGGCTG 24
|||||
86522 AAACCAATGACCAACTGGGCTG 86545
|||||

RESULT 50
AC027584/c
```

LOCUS AC027584 144003 bp DNA linear PRI 25-NOV-2001  
 DEFINITION Homo sapiens chromosome 18, clone RP11-214L13, complete sequence.  
 ACCESSION AC027584  
 VERSION AC027584.5 GI:16596606  
 HTG.  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 144003)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 18, clone RP11-214L13  
 Unpublished  
 2 (bases 1 to 144003)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domno,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczkv,J.,  
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 144003)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehoczkv,J., Levine,R., Liu,G.,  
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 2, 2001 this sequence version replaced gi:14994164.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu

CONTACT: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L9024  
 Center clone name: 214\_L13  
 -----  
 FEATURES  
 source  
 1. 144003  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18"  
 /clone="RP11-214L13"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 230..289  
 /rpt\_family="L2"  
 repeat\_region  
 complement (344..691)  
 /rpt\_family="L1MBC"  
 repeat\_region  
 complement (736..1232)  
 /rpt\_family="L1MBC"  
 repeat\_region  
 complement (1287..2755)  
 /rpt\_family="L1PA5"  
 repeat\_region  
 complement (2756..3066)  
 /rpt\_family="AluYa5"  
 repeat\_region  
 complement (3067..6236)  
 /rpt\_family="L1PA5"  
 repeat\_region  
 6239..6342  
 /rpt\_family="CT-rich"  
 repeat\_region  
 complement (6365..6657)  
 /rpt\_family="AluX"  
 repeat\_region  
 complement (6673..8150)  
 /rpt\_family="L1MBC"  
 repeat\_region  
 complement (9977..10170)  
 /rpt\_family="MIR"  
 repeat\_region  
 10843..10945  
 /rpt\_family="MIR3"  
 repeat\_region  
 11285..11322  
 /rpt\_family="AT-rich"  
 repeat\_region  
 complement (11536..11722)  
 /rpt\_family="L1MCA4a"  
 repeat\_region  
 complement (11788..12154)  
 /rpt\_family="HAL1b"  
 repeat\_region  
 complement (12247..12933)  
 /rpt\_family="HAL1b"  
 repeat\_region  
 13591..13614  
 /rpt\_family="TA)n"  
 repeat\_region  
 complement (13875..14038)  
 /rpt\_family="L1MCA4a"  
 repeat\_region  
 14173..14230  
 /rpt\_family="MIR3"  
 repeat\_region  
 15625..15796  
 /rpt\_family="MERS7B"  
 repeat\_region  
 15828..15901  
 /rpt\_family="MERS7A"  
 repeat\_region  
 complement (15903..16193)  
 /rpt\_family="AluJb"  
 repeat\_region  
 17096..17153  
 /rpt\_family="MERS7A-int"  
 repeat\_region  
 complement (17154..17448)  
 /rpt\_family="AluSq"  
 repeat\_region  
 17449..17529  
 /rpt\_family="MERS7A-int"  
 repeat\_region  
 17690..18056  
 /rpt\_family="MERS7A-int"  
 repeat\_region  
 18197..18302  
 /rpt\_family="MERS34"  
 repeat\_region  
 18303..18368  
 /rpt\_family="MERS34"  
 repeat\_region  
 18588..19552  
 /rpt\_family="MERS7-int"  
 repeat\_region  
 19535..20187  
 /rpt\_family="MERS7-int"  
 repeat\_region  
 20586..20784

repeat\_region /rpt family="MER57-int"  
21307..22250  
repeat\_region /rpt family="MER57-int"  
22251..22605  
repeat\_region /rpt family="MER57A"  
23090..23121  
repeat\_region /rpt family="(TA)n"  
complement(23122..23420)  
repeat\_region /rpt family="AluJo"  
23857..24010  
repeat\_region /rpt family="MIR3"  
24147..24211  
repeat\_region /rpt family="MER31A"  
24917..24940  
repeat\_region /rpt family="AT\_rich"  
25532..25715  
repeat\_region /rpt family="MER20"  
26047..26085  
repeat\_region /rpt family="AT\_rich"  
complement(26320..26452)  
repeat\_region /rpt family="LINE"  
26603..26630  
repeat\_region /rpt family="AT\_rich"  
complement(27226..27530)  
repeat\_region /rpt family="AluSx"  
complement(28029..28085)  
repeat\_region /rpt family="LINE"  
complement(28107..29600)  
repeat\_region /rpt family="LMB3"  
complement(30124..30460)  
repeat\_region /rpt family="LMB3"  
complement(31116..31622)  
repeat\_region /rpt family="LMB3"  
32859..32895  
repeat\_region /rpt family="AT\_rich"  
33018..33201  
repeat\_region /rpt family="MER58A"  
34924..35065  
repeat\_region /rpt family="MIR"  
complement(35066..35294)  
repeat\_region /rpt family="MER3"  
35295..35383  
repeat\_region /rpt family="MIR"  
35556..35815  
repeat\_region /rpt family="MIR"  
complement(36703..36851)  
repeat\_region /rpt family="L2"  
37033..37062  
repeat\_region /rpt family="AT\_rich"  
37142..37749  
repeat\_region /rpt family="L1PA15"  
37792..38085  
repeat\_region /rpt family="AluJo"  
39160..39343

Query Match 76.8%; Score 19.2; DB 8; Length 144003;  
Best Local Similarity 87.5%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 48295 AAAGCANTGACAACTGGGCTGT 48272

Search completed: February 3, 2006, 21:26:14  
Job time : 877 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-5  
Perfect score: 25  
Sequence: 1 aaaaacaatgacccaactgggtgt 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : N Geneseq\_21:  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20.8	83.2	57248	6	ABK83563	ABK83563 Human c
2	19.2	76.8	600	4	AAK11812	AAK11812 Human b
3	19.2	76.8	175603	12	ADQ97554	Adq97554 Mouse
C 4	18.8	75.2	6596	4	ABL08072	AbI08072 Drosoph
5	18.6	74.4	290547	13	ABD32598	ABd32598 Human
C 6	18.4	73.6	308	6	ABL62734	AbI62734 Breast
C 7	18.4	73.6	2146	12	ADQ22079	Adq22079 Human
8	18.2	72.8	650	13	ACN62534	Acn62534 Cotton
9	18.2	72.8	1003	8	ACA49099	Aca49099 Prokary
10	18.2	72.8	1047	8	ACA51846	Aca51846 Prokary
11	18.2	72.8	2319	13	ADX09339	Adx09339 Plant
12	18.2	72.8	54810	13	ABD32759	ABd32759 Mouse
C 13	18.2	72.8	66479	6	ABQ80567	Abq80567 Mutant
C 14	18.2	72.8	66479	6	ABQ80566	Abq80566 Mutant
C 15	18.2	72.8	66479	6	ABQ80568	Abq80568 Mutant
C 16	18.2	72.8	66479	6	ABQ80565	Abq80565 Human i
17	18.2	72.8	110000	13	ABD32804_4	Continuation (5
C 18	18	72.0	1522	6	ABZ16013	AbZ16013 Arabid
C 19	18	72.0	1522	8	ADA69122	Ada69122 Arabid

93	17	68.0	441	5	ABV19324	Human pro
94	17	68.0	484	10	ABV5553	Toxicity-
95	17	68.0	492	5	ABV49098	Human pro
c 96	17	68.0	500	11	ADM65753	NRY polym
c 97	17	68.0	501	6	ABV53676	Plant ful
c 98	17	68.0	519	13	ADK1946	Human pro
c 99	17	68.0	613	5	ABV55948	Human pro
c 100	17	68.0	700	4	AHL92872	Human Inf
c 101	17	68.0	728	4	ABL20019	Drosophil
c 102	17	68.0	761	12	ADO62623	Transcrip
c 103	17	68.0	765	12	ADQ17998	Human sof
c 104	17	68.0	984	6	ABO69809	Human sof
c 105	17	68.0	984	6	ABO68114	Listeria
c 106	17	68.0	984	8	ACA36535	Prokaryot
c 107	17	68.0	996	6	ABQ67967	Listeria
c 108	17	68.0	996	6	ABQ69994	Listeria
c 109	17	68.0	1077	14	ADW81848	MAP3K9 ma
c 110	17	68.0	1082	5	ABV28252	Human pro
c 111	17	68.0	1082	5	ABV22438	Human pro
c 112	17	68.0	1176	8	ACA33726	Prokaryot
c 113	17	68.0	1380	4	AAF44674	Novel pro
c 114	17	68.0	1380	12	ADI29372	Mouse MAR
c 115	17	68.0	1674	13	ADT19380	Plant cDN
c 116	17	68.0	2000	7	ADZ74598	Arabidops
c 117	17	68.0	2015	6	ABK63485	Rat seque
c 118	17	68.0	2015	10	ADB57938	Toxicity-
c 119	17	68.0	2015	11	ADW21828	Rat hepat
c 120	17	68.0	2023	11	ADW22482	Rat hepat
c 121	17	68.0	2026	10	ADB58195	Toxicity-
c 122	17	68.0	2026	10	ADB52699	Primary r
c 123	17	68.0	2026	10	ABT41903	Toxicity
c 124	17	68.0	2397	6	ABX13908	cDNA enco
c 125	17	68.0	3144	13	ADR07476	Full leng
c 126	17	68.0	3260	11	ADM01957	Human cDN
c 127	17	68.0	3377	4	ABL20018	Drosophil
c 128	17	68.0	3818	4	ABL18574	Drosophil
c 129	17	68.0	5508	10	ADE85828	Human pai
c 130	17	68.0	5508	10	ADE85827	Human pai
c 131	17	68.0	5508	10	ADE85824	Human pai
c 132	17	68.0	5508	10	ADE85826	Human pai
c 133	17	68.0	5775	6	ABQ70927	Listeria
c 134	17	68.0	8189	4	ABL01898	Drosophil
c 135	17	68.0	8497	14	AEA61764	Streptoco
c 136	17	68.0	9905	6	AAK98324	Human pur
c 137	17	68.0	9985	4	ABL05688	Drosophil
c 138	17	68.0	24888	13	ABD33500	Human can
c 139	17	68.0	28854	8	ABX95685	Human gen
c 140	17	68.0	33780	4	AAH24652	Nucleotid
c 141	17	68.0	40491	11	ACN44798	Human gen
c 142	17	68.0	42324	4	AAK84724	Human imm
c 143	17	68.0	104062	11	ACN44458	Human gen
c 144	17	68.0	110000	6	ABA03041_00	Listeria
c 145	17	68.0	110000	6	ABA03041_14	Continuation (15 o
c 146	17	68.0	117750	13	ABD32653	Human can
c 147	17	68.0	165554	11	ACN44108	Mouse gen
c 148	17	68.0	175378	14	ADV09445	Human ORP
c 149	17	68.0	256493	11	ACN44514	Human gen
c 150	16.8	67.2	885	4	AAH24195	Mouse age

ALIGNMENTS

RESULT 1  
ABK83563  
ID ABK83563 standard; cDNA; 57248 BP.  
XX  
AC ABK83563;

14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #134.

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
OS Homo sapiens.  
XX WO200228999-A2.  
XX 11-APR-2002.  
XX 03-OCT-2001; 2001WO-US030821.  
XX 03-OCT-2000; 2000US-0237189P.  
XX (GENE-) GENE LOGIC INC.  
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX WPI; 2002-435328/46.  
XX Detecting granulocyte activation by detecting differential expression of  
XX genes associated with granulocyte activation, which serves as diagnostic  
XX markers that is useful for monitoring disease states and drug toxicity.  
XX Claim 1; SEQ ID NO 134; 114pp; English.  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
XX DNA chip analysis as given in the specification, and comparing the  
XX expression level to an expression level in an unactivated GC, where  
XX differential expression of Gs is indicative of GCA. Also included are  
XX modulating (M2) GCA by contacting GC with an agent that alters the  
XX expression of at least one gene in Gs; (2) screening (M3) for an agent  
XX capable of modulating GCA or an inflammation (especially chronic) in a  
XX tissue, an allergic response in a subject, exposure of a subject to a  
XX pathogen or sterile inflammatory disease using the gene expression  
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a  
XX tissue, an allergic response in a subject, exposure of a subject to a  
XX pathogen or sterile inflammatory disease, by detecting the level of  
XX expression in a sample of the tissue of gene(s) from Gs, where the level  
XX of expression of the gene is indicative of inflammation; (4) treating  
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic  
XX response in a subject, exposure of a subject to a pathogen or sterile  
XX inflammatory disease, by contacting a tissue having inflammation with an  
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1  
XX is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful  
XX for screening an agent capable of modulating GCA preferably in an  
XX inflammation in a tissue; M4 is useful for detecting an inflammation  
XX (especially chronic) in a tissue, an allergic response in a subject,  
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
XX respiratory distress syndrome, inflammatory bowel disease, Crohn's  
XX disease, ulcerative colitis, periodontal disease; also bacterial  
XX infection, viral infection, parasitic infection, protozoal infection,  
XX fungal infection and M5 is useful for treating one of the above  
XX conditions. The present sequence represents a gene differentially  
XX expressed in granulocytes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 57248 BP; 15003 A; 13601 C; 13307 G; 15337 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 57248;  
Best Local Similarity 91.7%; Pred. No. 37;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 AAAAAACAATGACCAACTGGGCTG 24
Db 42908 AAAAAACAATGTCCAACTGGTCTG 42931
|||||

RESULT 2
AAK11812
ID AAK11812 standard; DNA; 600 BP.
XX
AC AAK11812;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 11803.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 11803; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 600 BP; 187 A; 104 C; 123 G; 186 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 4; Length 600;
Best Local Similarity 87.5%; Pred. No. 98;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCTG 25
Db 370 AAAAAACCATGACCAACTGGTTGT 393
|||||

RESULT 3
ADQ97554
ID ADQ97554 standard; DNA; 175603 BP.
XX
AC ADQ97554;
XX
DT 07-OCT-2004 (first entry)
XX
```

```
DE Mouse cancer associated sequence MD09-013, SEQ ID 531.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS Mus musculus.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
XX WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 531; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 175603 BP; 48715 A; 34685 C; 35401 G; 51161 T; 0 U; 5641 Other;

Query Match 76.8%; Score 19.2; DB 12; Length 175603;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCTGT 25
Db 147030 AAAAAACAATGACCAAGCGGTTGT 147053
|||||

RESULT 4
ABL08072/c
ID ABL08072 standard; cDNA; 6596 BP.
XX
AC ABL08072;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18698.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR
```

DR P-PSDB; ABB63969.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
PS Claim 1; SEQ ID NO 18698; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 6596 BP; 1724 A; 1400 C; 1614 G; 1858 T; 0 U; 0 Other;  
Query Match 75.2%; Score 18.8; DB 4; Length 6596;  
Best Local Similarity 90.9%; Pred. No. 2.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAAATGACCAACTGGGCT 23  
DB 6184 AAAAAAATGACCAACTGGACT 6163  
RESULT 5  
ABD32598  
ID ABD32598 standard; DNA; 290547 BP.  
XX  
AC ABD32598;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human cancer-associated genomic DNA HD7-220.  
XX  
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;  
KW leukaemia; lymphoma; CAP.  
XX  
OS Homo sapiens.  
XX  
PN WO2004074320-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 17-FEB-2004; 2004WO-US004730.  
XX  
PR 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-00388838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
PI Morris DW, Morris DW, Malandro MS;  
XX  
XX WPI; 2004-652914/63.  
XX  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX  
PS claim 16; seqid 103; 310pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-

CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer,  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 290547 BP; 78477 A; 46715 C; 46576 G; 86851 T; 0 U; 31928 Other;  
Query Match 74.4%; Score 18.6; DB 13; Length 290547;  
Best Local Similarity 84.0%; Pred. No. 4.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACAATGACCAACTGGGCTGT 25  
DB 26055 AAAAAAATATCAACAACACTGGGCTGT 26079  
RESULT 6  
ABL62734/C  
ID ABL62734 standard; DNA; 308 BP.  
XX  
AC ABL62734;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Breast cancer related gene sequence SEQ ID NO:1071.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US010838.  
XX  
PR 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.

25-SEP-2000; 2000US-0234923P.  
25-SEP-2000; 2000US-0234924P.  
25-SEP-2000; 2000US-0235077P.  
25-SEP-2000; 2000US-0235082P.  
25-SEP-2000; 2000US-0235134P.  
25-SEP-2000; 2000US-0235280P.  
26-SEP-2000; 2000US-0235637P.  
26-SEP-2000; 2000US-0235638P.  
27-SEP-2000; 2000US-0235711P.  
27-SEP-2000; 2000US-0235720P.  
27-SEP-2000; 2000US-0235840P.  
27-SEP-2000; 2000US-0235863P.  
28-SEP-2000; 2000US-0236028P.  
28-SEP-2000; 2000US-0236032P.  
28-SEP-2000; 2000US-0236033P.  
28-SEP-2000; 2000US-0236034P.  
28-SEP-2000; 2000US-0236109P.  
28-SEP-2000; 2000US-0236111P.  
29-SEP-2000; 2000US-0236842P.  
29-SEP-2000; 2000US-0236891P.  
02-OCT-2000; 2000US-0237172P.  
02-OCT-2000; 2000US-0237173P.  
02-OCT-2000; 2000US-0237278P.  
02-OCT-2000; 2000US-0237294P.  
02-OCT-2000; 2000US-0237295P.  
02-OCT-2000; 2000US-0237316P.  
03-OCT-2000; 2000US-0237425P.  
03-OCT-2000; 2000US-0237598P.  
03-OCT-2000; 2000US-0237604P.  
03-OCT-2000; 2000US-0237606P.  
03-OCT-2000; 2000US-0237608P.  
01-NOV-2000; 2000US-0244867P.  
01-NOV-2000; 2000US-0245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
PI WPI; 2002-188264/24.  
DR  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX  
XX Claim 1; SEQ ID NO 1071; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'  
CC tumour  
XX  
SQ Sequence 308 BP; 71 A; 64 C; 67 G; 106 T; 0 U; 0 Other;  
XX  
Query Match 73.6%; Score 18.4; DB 6; Length 308;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAACAATGTCACCACTGG 20  
|||||

Db 21 AAAAAACAATGTCACCACTGG 2  
RESULT 7  
ADQ22079/c  
ID ADQ22079 standard; DNA; 2146 BP.  
XX  
AC ADQ22079;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 4899.  
XX  
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
ds.  
XX  
OS Homo sapiens.  
XX  
FN WO2004048938-A2.  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnik A;  
XX  
XX WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
XX Example 2; SEQ ID NO 4899; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 2146 BP; 536 A; 443 C; 428 G; 739 T; 0 U; 0 Other;  
XX  
Query Match 73.6%; Score 18.4; DB 12; Length 2146;  
Best Local Similarity 95.0%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAACAATGTCACCACTGG 20  
|||||  
Db 679 AAAAAACAATGTCACCACTGG 660  
RESULT 8  
ACN62534  
ID ACN62534 standard; cDNA; 650 BP.  
XX  
AC ACN62534;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Cotton developing fibre EST Clone ID: L1B3830-001-Q1-K6-E8, SEQ:17315.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant;  
KW developing fibre; variety Nutcoton33B; library LIB3830; molecular tag;  
KW molecular marker; genetic mapping; molecular mapping; seed germination;  
KW plant growth; plant quality; plant yield; plant breeding;  
KW tissue printing; ss.  
XX  
OS Gossypium hirsutum.  
XX  
XX US2004123340-A1.  
XX  
XX 24-JUN-2004.  
XX  
XX 12-DEC-2001; 2001US-00021323.  
XX  
XX 14-DEC-2000; 2000US-0255619P.  
XX  
XX (DEIK/) DEIKMAN J.  
PA (FENG/) FENG P C C.  
PA (FING/) FINCHER K L.  
PA (ZIEG/) ZIEGLER T E.  
XX  
XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;  
PI  
XX  
XX WPI; 2004-479808/45.  
XX  
XX New isolated nucleic acid molecule that encodes a plant protein or its  
PT fragment, useful for isolating a variety of agronomically significant  
PT genes associated with plant growth, quality or yield, and as molecular  
PT tags to map genes.  
XX  
XX Claim 1; SEQ ID NO 17315; 34pp; English.  
XX  
XX The invention relates to 17880 cotton expressed sequence tags (ESTs;  
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated  
CC from primed or non-primed seeds from variety DP50B, mature seeds from  
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium  
CC tissue, developing fibres, carpel walls and septa from variety  
CC Nutcoton33B. The invention also relates to substantially purified  
CC proteins or their fragments encoded by nucleic acid molecules of the  
CC invention, and to transformed plants having a nucleic acid construct  
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as  
CC molecular tags to isolate genetic regions, to isolate genes, to map  
CC genes, to determine gene function, and to determine whether genes are  
CC members of a particular gene family. The nucleic acid molecules may be  
CC used for isolating a variety of agronomically significant genes  
CC associated with plant growth, quality, yield, and could also serve as  
CC links in metabolic and catabolic pathways. The nucleic acid molecules are  
CC also useful for identifying genes important in initiating and maintaining  
CC seed germination or that may be used to mitigate stresses encountered  
CC during seed germination. The ESTs additionally enable the acquisition of  
CC promoters and cis-regulatory elements which will be useful to express  
CC agronomically significant genes in these tissues and/or other tissues,  
CC and also permits the acquisition of molecular markers useful in breeding  
CC schemes, genetic and molecular mapping, and in cloning of agronomically  
CC significant genes. The nucleic acid molecules are further useful for  
CC detecting the expression level or pattern of a protein or mRNA and for  
CC detecting the presence or quantity of a protein by tissue printing. The  
CC present sequence represents a specifically claimed EST isolated from a  
CC cotton variety Nutcoton33B developing fibre cDNA library (LIB3830). The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the US  
CC patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340  
XX  
SQ Sequence 650 BP; 193 A; 119 C; 136 G; 201 T; 0 U; 1 Other;  
Query Match 72.8%; Score 18.2; DB 13; Length 650;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 2 AAAACAATGACCAACTGGGCTG 24  
DB 18 AAAACAATGACGACAGAGCTG 40

## RESULT 9

ACA49099

ID ACA49099 standard; DNA; 1003 BP.

XX

AC ACA49099;

XX

DT 19-JUN-2003 (first entry)

XX

DE Prokaryotic essential gene #30756.

XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX

OS Salmonella paratyphi.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US0009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR

PR 06-SEP-2001; 2001US-00948993.

PR

PR 25-OCT-2001; 2001US-0342923P.

PR

PR 08-FEB-2002; 2002US-00072851.

PR

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

XX

P-PSDB; ABU45229.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 14; SEQ ID NO 36969; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway;

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1003 BP; 280 A; 249 C; 264 G; 210 T; 0 U; 0 Other;  
Query Match 72.8%; Score 18.2; DB 8; Length 1003;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACAATGACCAACTGGGCTG 24  
||||| ||||| ||||| ||||| |||||  
Db 399 AAAAACCATTACCAGCTGGGCTG 421  
RESULT 10  
ACA51846  
ID ACA51846 standard; DNA; 1047 BP.  
XX  
AC ACA51846;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #33503.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Salmomella typhi.  
XX  
PN WO20027183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PP 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362899P.  
XX  
PA (BLIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU47976.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 39716; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
to which each of the strains is present in a culture or collection of  
strains; or (13) identifying the target of a compound that inhibits the  
proliferation of an organism. The antisense nucleic acids are useful for  
identifying proteins or screening for homologous nucleic acids required  
for cellular proliferation to isolate candidate molecules for rational  
drug discovery programs, or for screening homologous nucleic acids  
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
prokaryotic essential genes. Note: The sequence data for this patent did  
not form part of the printed specification, but was obtained in an  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1047 BP; 295 A; 261 C; 273 G; 218 T; 0 U; 0 Other;  
Query Match 72.8%; Score 18.2; DB 8; Length 1047;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACAATGACCAACTGGGCTG 24  
||||| ||||| ||||| ||||| |||||  
Db 435 AAAAACCATTACCAGCTGGGCTG 457  
RESULT 11  
ADX09399  
ID ADX09399 standard; cDNA; 2319 BP.  
XX  
AC ADX09399;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 3974.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactonnanan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX  
XX Unidentified.  
XX  
PN US2004034888-A1.  
XX  
PD 19-FEB-2004.  
XX  
XX 28-APR-2003; 2003US-00425114.  
XX  
XX 06-MAY-1999; 99US-00304517.  
XX  
XX 05-NOV-2001; 2001US-00985678.  
XX  
XX (LIU/) LIU J.  
XX (ZHOU/) ZHOU Y.  
XX (KOVA/) KOVALIC D K.  
XX (SCRE/) SCREEN S E.  
XX (TABAS/) TABASKA J E.  
XX (CAOY/) CAO Y.  
XX  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
PI WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
XX Claim 1; SEQ ID NO 3974; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a



CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomanan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 2319 BP; 616 A; 487 C; 602 G; 614 T; 0 U; 0 Other;  
  
Query Match 72.8%; Score 18.2; DB 13; Length 2319;  
Best Local Similarity 87.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAACAATGACCAACTGGGCT 23  
|||||  
Db 2012 AAAAAACAATGACCAACTGGGCT 2034  
  
RESULT 12  
ABD32759/c  
ID ABD32759 standard; DNA; 54810 BP.  
XX  
AC ABD32759;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Mouse cancer-associated genomic DNA MD16-030.  
XX  
KW Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;  
XX leukaemia; lymphoma; CAP.  
XX  
OS Mus musculus.  
XX  
PN WO2004074320-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 17-FEB-2004; 2004WO-US004730.  
XX  
PR 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-00388838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
PI Morris DW, Morris DW, Malandro MS;  
XX  
XX WPI; 2004-652914/63.  
XX  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX  
XX disclosure; seqid 365; 310pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising

CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 54810 BP; 14231 A; 10582 C; 11156 G; 18235 T; 0 U; 606 Other;  
  
Query Match 72.8%; Score 18.2; DB 13; Length 54810;  
Best Local Similarity 87.0%; Pred. No. 5.7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 AAAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 42877 ATAAACAATCAACCAACTGGGATGT 42855  
  
RESULT 13  
ABQ80567/c  
ID ABQ80567 standard; DNA; 66479 BP.  
XX  
AC ABQ80567;  
XX  
DT 08-NOV-2002 (first entry)  
XX  
DE Mutant human IKBKAP gene #2.  
XX  
KW Human; IKBKAP; Familial Dysautonomia; FD; Riley-Day syndrome;  
KW Hereditary Sensory and Autonomic Neuropathy Type III; carrier screening;  
XX PD2; mutation; gene; chromosome 9q31; ds.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT mutation replace(33714,G)  
FT /\*tag= a  
XX  
XX WO200259381-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 07-JAN-2002; 2002WO-US0000473.  
XX  
PR 06-JAN-2001; 2001US-0260080P.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
XX Slaugenhaupt S, Gusella JF;  
XX

DR WPI; 2002-674806/72.  
 XX New IKBKAP genes with mutations, useful for identifying a subject with  
 PT familial dysautonomia (FD), or for rapid carrier screening in the  
 PT Ashkenazi Jewish population, e.g. screening presymptomatic homozygotes or  
 PT prenatal diagnosis.  
 XX  
 XX Claim 1; Page; 109pp; English.  
 XX  
 CC The present invention relates to methods and compositions useful for  
 CC detecting mutations which cause Familial Dysautonomia (FD, Riley-Day  
 CC syndrome, Hereditary Sensory and Autonomic Neuropathy Type III) [OMIM  
 CC 223900]. It was found that mutations in the IKBKAP gene (see ABQ80565)  
 CC are associated with FD. The mutation associated with the major haplotype  
 CC of FD, FD1 mutation, is a base pair (bp) mutation, where the thymine  
 CC nucleotide located at bp 6 of intron 20 in the IKBKAP gene is replaced  
 CC with a cytosine. This results in skipping of exon 20 in the mRNA from FD  
 CC patients, although they continue to express varying levels of wild-type  
 CC message in a tissue-specific manner. The mutation associated with the  
 CC minor haplotype, FD2 mutation, is a bp mutation, where the guanine  
 CC nucleotide at bp 2397 (bp 73 of exon 19) is replaced with a cytosine.  
 CC This bp mutation causes an arginine to proline missense mutation (R696P)  
 CC in the IKBKAP protein, which is predicted to disrupt a potential  
 CC phosphorylation site. The IKBKAP nucleic acid sequences are useful for  
 CC identifying a subject with FD and for rapid carrier screening. The IKBKAP  
 CC gene contains 37 exons and maps to chromosome 9q31. Note: the present  
 CC sequence was not shown in the specification, but was derived from the  
 CC human wild-type IKBKAP sequence given in Fig 6  
 XX  
 XX Sequence 66479 BP; 18271 A; 12399 C; 14128 G; 21681 T; 0 U; 0 Other;  
 SQ  
 Query Match 72.8%; Score 18.2; DB 6; Length 66479;  
 Best Local Similarity 87.0%; Pred. No. 5.9e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAACAATGACCAACTGGGCT 23  
 DB 60341 AATAAACAAATGACCAATGGCCT 60319  
 RESULT 14  
 ABQ80566/c  
 ID ABQ80566 standard; DNA; 66479 BP.  
 XX  
 AC ABQ80566;  
 XX  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE Mutant human IKBKAP gene #1.  
 XX  
 KW Human; IKBKAP; Familial Dysautonomia; FD; Riley-Day syndrome;  
 KW Hereditary Sensory and Autonomic Neuropathy Type III; carrier screening;  
 KW FD1; mutation; gene; chromosome 9q31; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT mutation replace(34201,T)  
 FT /\*tag= a  
 FT mutation /\*tag= b  
 XX  
 PN WO200259381-A2.  
 XX  
 XX 01-AUG-2002.  
 XX  
 XX 07-JAN-2002; 2002WO-US0000473.  
 XX  
 XX 06-JAN-2001; 2001US-0260080P.  
 XX (GEO ) GEN HOSPITAL CORP.  
 XX  
 XX Slagenhaupt S, Gusella JF;

DR WPI; 2002-674806/72.  
 XX New IKBKAP genes with mutations, useful for identifying a subject with  
 PT familial dysautonomia (FD), or for rapid carrier screening in the  
 PT Ashkenazi Jewish population, e.g. screening presymptomatic homozygotes or  
 PT prenatal diagnosis.  
 XX  
 XX Claim 1; Page; 109pp; English.  
 XX  
 CC The present invention relates to methods and compositions useful for  
 CC detecting mutations which cause Familial Dysautonomia (FD, Riley-Day  
 CC syndrome, Hereditary Sensory and Autonomic Neuropathy Type III) [OMIM  
 CC 223900]. It was found that mutations in the IKBKAP gene (see ABQ80565)  
 CC are associated with FD. The mutation associated with the major haplotype  
 CC of FD, FD1 mutation, is a base pair (bp) mutation, where the thymine  
 CC nucleotide located at bp 6 of intron 20 in the IKBKAP gene is replaced  
 CC with a cytosine. This results in skipping of exon 20 in the mRNA from FD  
 CC patients, although they continue to express varying levels of wild-type  
 CC message in a tissue-specific manner. The mutation associated with the  
 CC minor haplotype, FD2 mutation, is a bp mutation, where the guanine  
 CC nucleotide at bp 2397 (bp 73 of exon 19) is replaced with a cytosine.  
 CC This bp mutation causes an arginine to proline missense mutation (R696P)  
 CC in the IKBKAP protein, which is predicted to disrupt a potential  
 CC phosphorylation site. The IKBKAP nucleic acid sequences are useful for  
 CC identifying a subject with FD and for rapid carrier screening. The IKBKAP  
 CC gene contains 37 exons and maps to chromosome 9q31. Note: the present  
 CC sequence was not shown in the specification, but was derived from the  
 CC human wild-type IKBKAP sequence given in Fig 6  
 XX  
 XX Sequence 66479 BP; 18271 A; 12399 C; 14129 G; 21680 T; 0 U; 0 Other;  
 SQ  
 Query Match 72.8%; Score 18.2; DB 6; Length 66479;  
 Best Local Similarity 87.0%; Pred. No. 5.9e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAACAATGACCAACTGGGCT 23  
 DB 60341 AATAAACAAATGACCAATGGCCT 60319  
 RESULT 15  
 ABQ80568/c  
 ID ABQ80568 standard; DNA; 66479 BP.  
 XX  
 AC ABQ80568;  
 XX  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE Mutant human IKBKAP gene #3.  
 XX  
 KW Human; IKBKAP; Familial Dysautonomia; FD; Riley-Day syndrome;  
 KW Hereditary Sensory and Autonomic Neuropathy Type III; carrier screening;  
 KW FD1; mutation; gene; chromosome 9q31; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT mutation replace(33714,G)  
 FT /\*tag= a  
 FT mutation replace(34201,T)  
 FT /\*tag= b  
 XX  
 PN WO200259381-A2.  
 XX  
 XX 01-AUG-2002.  
 XX  
 XX 07-JAN-2002; 2002WO-US0000473.  
 XX  
 XX 06-JAN-2001; 2001US-0260080P.  
 XX (GEO ) GEN HOSPITAL CORP.

PI Slangenhaupt S, Gusella JF;  
XX WPI; 2002-674806/72.  
XX  
PT New IKBKAP genes with mutations, useful for identifying a subject with  
PT familial dysautonomia (FD), or for rapid carrier screening in the  
PT Ashkenazi Jewish population, e.g. screening presymptomatic homozygotes or  
PT prenatal diagnosis.  
XX Claim 1; Page; 109pp; English.  
XX  
CC The present invention relates to methods and compositions useful for  
CC detecting mutations which cause Familial Dysautonomia (FD, Riley-Day  
CC syndrome, Hereditary Sensory and Autonomic Neuropathy Type III) [OMIM  
CC 223900]. It was found that mutations in the IKBKAP gene (see AB080565)  
CC are associated with FD. The mutation associated with the major haplotype  
CC of FD, FDI mutation, is a base pair (bp) mutation, where the thymine  
CC nucleotide located at bp 6 of intron 20 in the IKBKAP gene is replaced  
CC with a cytosine. This results in skipping of exon 20 in the mRNA from FD  
CC patients, although they continue to express varying levels of wild-type  
CC message in a tissue-specific manner. The mutation associated with the  
CC minor haplotype, FD2 mutation, is a bp mutation, where the guanine  
CC nucleotide at bp 2397 (bp 73 of exon 19) is replaced with a cytosine.  
CC This bp mutation causes an arginine to proline missense mutation (R696P)  
CC in the IKBKAP protein, which is predicted to disrupt a potential  
CC phosphorylation site. The IKBKAP nucleic acid sequences are useful for  
CC identifying a subject with FD and for rapid carrier screening. The IKBKAP  
CC gene contains 37 exons and maps to chromosome 9q31. Note: the present  
CC sequence was not shown in the specification, but was derived from the  
CC human wild-type IKBKAP sequence given in Fig 6  
XX  
SQ Sequence 66479 BP; 18271 A; 12400 C; 14128 G; 21680 T; 0 U; 0 Other;  
Query Match 72.8%; Score 18.2; DB 6; Length 66479;  
Best Local Similarity 87.0%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACATGACCAACTGGGCT 23  
DB 60341 AATAAACATGACCAATGGCCT 60319  
RESULT 16  
AB080565/C  
ID ABQ80565 standard; DNA; 66479 BP.  
XX  
AC ABQ80565;  
XX  
DT 08-NOV-2002 (first entry)  
XX  
DE Human IKBKAP wild-type gene.  
XX  
KW Human; IKBKAP; Familial Dysautonomia; FD; Riley-Day syndrome;  
KW Hereditary Sensory and Autonomic Neuropathy Type III; carrier screening;  
KW gene; chromosome 9q31; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200259381-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 07-JAN-2002; 2002WO-US000473.  
XX  
PR 06-JAN-2001; 2001US-0260080P.  
XX  
PA (GEO ) GEN HOSPITAL CORP.  
XX  
PI Slangenhaupt S, Gusella JF;  
XX WPI; 2002-674806/72.  
XX  
PT New IKBKAP genes with mutations, useful for identifying a subject with

PT familial dysautonomia (FD), or for rapid carrier screening in the  
PT Ashkenazi Jewish population, e.g. screening presymptomatic homozygotes or  
PT prenatal diagnosis.  
XX Claim 1; Fig 6; 109pp; English.  
XX  
CC The present invention relates to methods and compositions useful for  
CC detecting mutations which cause Familial Dysautonomia (FD, Riley-Day  
CC syndrome, Hereditary Sensory and Autonomic Neuropathy Type III) [OMIM  
CC 223900]. It was found that mutations in the IKBKAP gene (the present  
CC sequence) are associated with FD. The mutation associated with the major  
CC haplotype of FD, FDI mutation, is a base pair (bp) mutation, where the  
CC thymine nucleotide located at bp 6 of intron 20 in the IKBKAP gene is  
CC replaced with a cytosine. This results in skipping of exon 20 in the mRNA  
CC from FD patients, although they continue to express varying levels of  
CC wild-type message in a tissue-specific manner. The mutation associated  
CC with the minor haplotype, FD2 mutation, is a bp mutation, where the  
CC guanine nucleotide at bp 2397 (bp 73 of exon 19) is replaced with a  
CC cytosine. This bp mutation causes an arginine to proline missense  
CC mutation (R696P) in the IKBKAP protein, which is predicted to disrupt a  
CC potential phosphorylation site. The IKBKAP nucleic acid sequences are  
CC useful for identifying a subject with FD and for rapid carrier screening.  
CC The IKBKAP gene contains 37 exons and maps to chromosome 9q31  
XX  
SQ Sequence 66479 BP; 18271 A; 12398 C; 14129 G; 21681 T; 0 U; 0 Other;  
Query Match 72.8%; Score 18.2; DB 6; Length 66479;  
Best Local Similarity 87.0%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACATGACCAACTGGGCT 23  
DB 60341 AATAAACATGACCAATGGCCT 60319  
RESULT 17  
ABD32804\_4  
Continuation (5 of 6) of ABD32804 from base 400001 (Mouse cancer-associated genomic DNA  
WP Sequence split into 6 fragments LOCUS ABD32804 Accession Abd32804  
WP Fragment Name Begin End  
WP ABD32804\_0 1 110000  
WP ABD32804\_1 100001 210000  
WP ABD32804\_2 200001 310000  
WP ABD32804\_3 300001 410000  
WP ABD32804\_4 400001 510000  
WP ABD32804\_5 500001 608916  
Query Match 72.8%; Score 18.2; DB 13; Length 110000;  
Best Local Similarity 87.0%; Pred. No. 6.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAAACATGACCAACTGGGCTG 24  
DB 32617 AACGCAATGACCAACTGGCTG 32639  
RESULT 18  
ABZ16013/C  
ID ABZ16013 standard; DNA; 1522 BP.  
XX  
AC ABZ16013;  
XX  
DT 21-JAN-2003 (first entry)  
XX  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 3818.  
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX  
PN WO200216655-A2.  
XX  
PD 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US026685.  
XX  
XX 24-AUG-2000; 2000US-0227866P.  
PR 26-JAN-2001; 2001US-0264647P.  
PR 22-JUN-2001; 2001US-0300111P.  
XX  
XX (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
XX  
XX Identifying a stress condition to which a plant cell has been exposed and  
PT producing plants with increased tolerance to these abiotic stresses.  
XX  
XX Claim 144; SEQ ID NO 3818; 577pp + Sequence Listing; English.  
XX  
XX The invention relates to identifying a stress condition to which a plant  
CC cell has been exposed, comprising: (a) contacting nucleic acid  
CC representative of expressed polynucleotides in the plant cell with an  
CC array or probes representative of the plant cell genome; and (b)  
CC detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants  
CC with increased tolerance to abiotic stress. The present sequence is that  
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used  
CC in methods of the invention. Note: The sequence data for this patent is  
CC not represented in the printed specification but is based on sequence  
CC information supplied to Derwent by the European Patent Office  
XX  
XX Sequence 1522 BP; 392 A; 246 C; 293 G; 591 T; 0 U; 0 Other;  
SQ  
  
Query Match 72.0%; Score 18; DB 6; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAACAATGACCAACT 18  
Db 1386 AAAAAACAATGACCAACT 1369  
  
RESULT 19  
ADA69122/c  
ID ADA69122 standard; DNA; 1522 BP.  
XX  
XX ADA69122;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX Arabidopsis thaliana gene, SEQ ID 2445.  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; gene; ds.  
XX  
XX Arabidopsis thaliana.  
XX  
XX WO2003000898-A1.  
XX  
XX 03-JAN-2003.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX Identifying at least one gene involved in plant resistance or response to

PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 27; SEQ ID NO 2445; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 1522 BP; 392 A; 246 C; 293 G; 591 T; 0 U; 0 Other;  
SQ  
  
Query Match 72.0%; Score 18; DB 8; Length 1522;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAACAATGACCAACT 18  
Db 1386 AAAAAACAATGACCAACT 1369  
  
RESULT 20  
ABZ36201/c  
ID ABZ36201 standard; cDNA; 3824 BP.  
XX  
XX ABZ36201;  
XX  
XX 10-FEB-2003 (first entry)  
XX  
XX Human secretory polynucleotide SPTM SEQ ID NO 365.  
XX  
XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;  
XX asthma; Crohn's disease; neurological disorder; epilepsy; cancer;  
XX Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;  
XX multiple sclerosis; Parkinson's disease; cell proliferative disorder;  
XX anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;  
XX neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;  
XX antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;  
XX secretory polynucleotide; secretory protein; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200283876-A2.  
XX  
XX 24-OCT-2002.  
XX  
XX 27-MAR-2002; 2002WO-US009921.  
XX  
XX 29-MAR-2001; 2001US-0280067P.  
XX 29-MAR-2001; 2001US-0280068P.  
XX 16-MAY-2001; 2001US-0291280P.  
XX 17-MAY-2001; 2001US-0291829P.  
XX 19-JUN-2001; 2001US-0291849P.  
XX 20-JUN-2001; 2001US-0299428P.  
XX 20-JUN-2001; 2001US-0299776P.  
XX 20-JUN-2001; 2001US-0300001P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;  
XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
XX Flores V, Marwaha R, Lo A, Lan RY, Urashka MB;  
XX WPI; 2003-075543/07.  
XX  
XX DR

DR P-PSDB; ABP75759.  
XX New human secretory proteins and polynucleotides, useful for diagnosing,  
PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),  
PT neurological disorders (e.g. Alzheimer's), or cell proliferations or  
PT cancers.  
XX Claim 1; SEQ ID NO 365; 459pp + Sequence Listing; English.  
XX  
XX The invention relates to a secretory polynucleotide (designated sptm)  
CC comprising any of 567 polynucleotide sequences (AB235837-AB236403), a  
CC naturally occurring polynucleotide sequence at least 90 % identical to  
CC the polynucleotide sequence, a polynucleotide complementary to them or an  
CC RNA equivalent of them. The polypeptide or polynucleotide are useful for  
CC treating, preventing or diagnosing a disease or condition associated with  
CC the expression of functional SPTM. These are particularly useful for  
CC diagnosing, treating or preventing autoimmune/inflammatory disorders  
CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's  
CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,  
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,  
CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,  
CC schizophrenia or amnesia), or cell proliferative disorders (e.g.  
CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,  
CC breast, cervix or prostate). Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3824 BP; 1005 A; 937 C; 934 G; 948 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 8; Length 3824;  
Best Local Similarity 90.5%; Pred. No. 5.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AACAAATGACCACTGGGCTGT 25  
Db 2977 AACAGGACCCACTGGGCTGT 2957  
RESULT 21  
ADE28224/c  
ID ADE28224 standard; DNA; 5821 BP.  
XX ADE28224;  
AC  
XX  
XX 29-JAN-2004 (first entry)  
DT  
XX Human MDDT DNA - SEQ ID 74.  
DE  
XX human; MDDT; molecule for disease detection and treatment; cytostatic;  
KW antiarteriosclerotic; anticonvulsant; nootropic; neuroprotective;  
KW cerebroprotective; anti-HIV; antiallergic; antiinflammatory;  
KW thymometric; cell proliferative; cancer; atherosclerosis; neurological;  
KW epilepsy; Huntington's disease; stroke; immune; inflammatory; AIDS;  
KW allergy; developmental disorder; hypothyroidism; Cushing's syndrome;  
KW infection; ds; gene.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003046152-A2.  
PN  
XX  
XX 05-JUN-2003.  
PD  
XX  
XX 25-NOV-2002; 2002WO-US038446.  
PF  
XX  
XX 28-NOV-2001; 2001US-0334182P.  
PR  
XX 18-DEC-2001; 2001US-0342052P.  
PR  
XX 18-JAN-2002; 2002US-0350410P.  
PR  
XX 01-FEB-2002; 2002US-0353284P.  
PR  
XX 08-MAR-2002; 2002US-0363649P.  
PR  
XX (INCY-) INCYTE GENOMICS INC.  
PA  
XX

PI Tang YT, Chawla NK, Lu DAM, Khan FA, Gandhi AR, Swarnakar A;  
PI Azimzai Y, Marquis JP, Sprague WW, Emerling BM, Yue H, Borowsky ML;  
PI Becha SD, Ison CH, Elliott VS, Hafalia AJA, Ring HZ, Warren BA;  
PI Gietzen KU, Tran UK, Lee SY, Lee EA, Richardson TW, Kable AE;  
PI Burford N, Lehr-Mason PM, Gorvad AE, Lee S, Blake JJ, Honchell CD;  
PI Thangavelu K, Ramkumar J, Chien D, Jin P, Chang H, Baughn MR;  
PI Nguyen DB, Khare R, Bhatia U, Burrill JD, Ho A, Zheng W;  
XX WPI; 2003-513643/48.  
DR P-PSDB; ADE28155.  
XX  
XX New human molecules for disease detection and treatment (MDDT), useful  
PT for diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant MDDT expression e.g. cancer, AIDS, atherosclerosis,  
PT epilepsy, or infections.  
XX  
XX Claim 5; SEQ ID NO 74; 314pp; English.  
XX  
XX The invention relates to a novel isolated human MDDT (molecule for  
CC disease detection and treatment) polypeptide. The polypeptide of the  
CC invention demonstrates cytostatic, antiarteriosclerotic, anticonvulsant,  
CC nootropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,  
CC antiinflammatory and thymometric activities and may be useful for  
CC diagnosing, treating and preventing a variety of diseases including cell  
CC proliferative diseases such as cancer and atherosclerosis, neurological  
CC diseases, in particular epilepsy, Huntington's disease and stroke, immune  
CC or inflammatory diseases including AIDS and allergies and developmental  
CC disorders including hypothyroidism and Cushing's syndrome, as well as  
CC infections. The current sequence is that of the human MDDT DNA of the  
CC invention.  
XX  
XX Sequence 5821 BP; 1439 A; 1569 C; 1553 G; 1260 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 10; Length 5821;  
Best Local Similarity 90.5%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AACAAATGACCACTGGGCTGT 25  
Db 5096 AACAGGACCCACTGGGCTGT 5076  
RESULT 22  
ADY19422/c  
ID ADY19422 standard; DNA; 6225 BP.  
XX ADY19422;  
AC  
XX  
XX 05-MAY-2005 (first entry)  
DT  
XX  
XX DNA encoding a PRO polypeptide, SEQ ID NO 5228.  
DE  
XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;  
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;  
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;  
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;  
KW Antiallergic; ds; gene; diagnosis.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2005016962-A2.  
PN  
XX  
XX 24-FEB-2005.  
PD  
XX  
XX 11-AUG-2004; 2004WO-US026249.  
PF  
XX  
XX 11-AUG-2003; 2003US-0493546P.  
PR  
XX (GETH) GENENTECH INC.  
PA  
XX Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;  
PI WPI; 2005-182330/19.  
DR

XX New nucleic acid encoding PRO polypeptide, useful for diagnosing and  
PT treating an immune related disorder, e.g. systemic lupus erythematosus,  
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.  
XX  
XX Claim 1; SEQ ID NO 5228; 159pp; English.  
XX  
XX The invention relates to an isolated nucleic acid encoding a PRO  
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,  
CC composition, and method are useful for diagnosing and treating an immune  
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid  
CC arthritis. The present sequence represents a PRO polypeptide.  
XX  
XX Sequence 6225 BP; 1629 A; 1534 C; 1562 G; 1500 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 14; Length 6225;  
Best Local Similarity 90.5%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AACAAATGACCACTGGGCTGT 25  
Db 4359 AACAGGACCCACTGGGCTGT 4339  
RESULT 23  
ABQ61046/c  
ID ABQ61046 standard; cDNA; 7028 BP.  
XX  
AC ABQ61046;  
XX  
XX 26-FEB-2003 (first entry)  
XX  
XX Membrane protein MEMAP-32 encoding sequence.  
XX  
XX Neuroprotective; immunomodulator; cancer; chromosome 11; cytostatic;  
KW anti-inflammatory; gene therapy; nutritional supplement; wound; burn;  
KW ulcer; Alzheimer's disease; Huntington's disease;  
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;  
KW vulnary; gene; SB.  
XX  
XX Homo sapiens.  
XX  
XX WO200231111-A2.  
XX  
XX 18-APR-2002.  
XX  
XX 11-OCT-2001; 2001WO-US027760.  
XX  
XX 12-OCT-2000; 2000US-00687527.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
XX WPI; 2002-426278/45.  
XX  
XX N-PSDB; ABP43802.  
XX  
XX New polypeptides and their encoded proteins, useful as nutritional  
PT sources or supplements, or in gene therapy, particularly for treating  
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
PT inflammation.  
XX  
XX Claim 1; SEQ ID # 259; 357pp + Sequence Listing; English.  
XX  
XX The invention relates to 446 newly isolated polynucleotide sequences. The  
CC activity of polynucleotides of the invention may be described as,  
CC inflammatory, neuroprotective, immunomodulator, cytostatic and anti-  
CC cancer. Compositions comprising nucleic acids of the invention are  
CC useful for treating a mammalian subject, or as nutritional sources or  
CC supplements. These are useful in gene therapy, particularly for treating  
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in  
CC diagnostic and research methods. The sequences given in records ABQ60788-  
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 7028 BP; 1786 A; 1790 C; 1749 G; 1703 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 6; Length 7028;  
Best Local Similarity 90.5%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AACAAATGACCACTGGGCTGT 25  
Db 4424 AACAGGACCCACTGGGCTGT 4404  
RESULT 24  
ADQ24033/c  
ID ADQ24033 standard; DNA; 7600 BP.  
XX  
XX AC ADQ24033;  
XX  
XX 26-AUG-2004 (first entry)  
XX  
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6853.  
XX  
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2004048938-A2.  
XX  
XX 10-JUN-2004.  
XX  
XX 26-NOV-2003; 2003WO-US038193.  
XX  
XX 26-NOV-2002; 2002US-0429739P.  
XX  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX  
XX Aziz N, Ginsburg WM, Zlotnik A;  
XX  
XX WPI; 2004-441208/41.  
XX  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
XX Example 2; SEQ ID NO 6853; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
XX Sequence 7600 BP; 1939 A; 1919 C; 1880 G; 1829 T; 0 U; 33 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 12; Length 7600;  
Best Local Similarity 90.5%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACATGACCACTGGGCTGT 25  
 DB 4359 AACAGGACCCACTGGGCTGT 4339

RESULT 25  
 ACL56900  
 ID ACL56900 standard; cDNA; 99 BP.  
 AC AC  
 XX ACL56900;  
 XX 24-MAR-2005 (first entry)  
 XX Human colon cancer differentially expressed polynucleotide, SEQ ID:3035.  
 DE  
 XX Differential expression; diagnosis; therapy; drug screening; cancer;  
 KW neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;  
 KW ss.  
 XX Homo sapiens.  
 OS  
 XX WO2005000087-A2.  
 PN  
 XX 06-JAN-2005.  
 PD  
 XX 13-MAY-2004; 2004WO-US015421.  
 PF  
 XX 03-JUN-2003; 2003US-0475872P.  
 PR  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Randazzo F, Moler E, Escobedo J, Garcia PD;  
 PI  
 XX WPI; 2005-075421/08.  
 DR  
 XX New isolated polynucleotides, which are differentially expressed in colon  
 PT cancer cell, useful for treating cancer, e.g. colon cancer, breast  
 PT cancer, or pancreatic cancer.  
 PT  
 XX Claim 1; SEQ ID NO 3035; 97pp; English.  
 PS  
 XX The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which  
 CC are differentially expressed in colon cancer cells. The invention also  
 CC relates to vectors and host cells comprising a differentially expressed  
 CC polynucleotide of the invention; a method for detecting a cancerous cell  
 CC by detection of a gene product of the polynucleotides; a method for  
 CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product  
 CC of the polynucleotides; a method of treating an individual with cancer by  
 CC administration of a modulator of a gene product of the polynucleotides;  
 CC and an isolated antibody that specifically binds to a polypeptide encoded  
 CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,  
 CC antibodies, and methods are useful for the detection of cancerous cells;  
 CC for the diagnosis, prognosis and management of cancer; for the  
 CC identification of agents that modulate the phenotype of cancerous cells;  
 CC for the identification of therapeutic targets for cancer chemotherapy;  
 CC and for the treatment of cancer, especially colon cancer and metastasized  
 CC colon cancer, but also breast or pancreatic cancer. The polynucleotides  
 CC are also useful as a source of probes or primers for use in diagnostic  
 CC methods. The differentially expressed polynucleotides or their encoded  
 CC proteins can additionally be used as vaccines to modulate primary immune  
 CC responses for the prevention or treatment of cancer. The present sequence  
 CC represents a specifically claimed polynucleotide which is differentially  
 CC expressed in colon cancer. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 99 BP; 16 A; 34 C; 17 G; 32 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 14; Length 99;  
 Best Local Similarity 83.3%; Pred. No. 4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACAATGACCAACTGGGCTGT 25  
 DB 27 AAATTCAATCACTGGGCTGT 50

RESULT 26  
 ABX35670/c  
 ID ABX35670 standard; cDNA; 397 BP.  
 XX AC  
 XX ABX35670;  
 XX 20-FEB-2003 (first entry)  
 XX Bovine EST associated with lactation/muscle/fat deposition #835.  
 DE  
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.  
 XX  
 OS Bos Taurus.  
 OS  
 XX US2002137139-A1.  
 PN  
 XX 26-SEP-2002.  
 PD  
 XX 24-SEP-2001; 2001US-00960352.  
 PF  
 XX 12-JAN-1999; 99US-0115707P.  
 PR  
 XX 11-JAN-2000; 2000US-00480902.  
 PR  
 XX (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
 PI  
 XX WPI; 2003-110599/10.  
 DR  
 XX New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and analysis,  
 PT cattle breeding, or for genetically improving cattle.  
 PT  
 XX Claim 2; SEQ ID NO 835; 245pp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived from  
 CC cattle, and the LMFD nucleic acid can specifically hybridize to a second  
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,  
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are  
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic  
 CC acid linked to a promoter and a 3' non-translated sequence that  
 CC functions in the cell to cause termination of transcription and addition  
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
 CC (2) determining a level or pattern of a molecule in a bovine cell or  
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a  
 CC complementary nucleic acid molecule obtained from the bovine cell or  
 CC tissue, where hybridisation between the marker nucleic acid and the  
 CC complementary nucleic acid permits the detection of the molecule; and (b)  
 CC detecting the level or pattern of the complementary nucleic acid, where  
 CC the detection of the complementary nucleic acid is predictive of the  
 CC level or pattern of the molecule. The LMFD nucleic acid is used for  
 CC determining a level or pattern of a molecule in a bovine cell or tissue.  
 CC It is useful for genome mapping, gene identification and analysis, cattle  
 CC breeding, preparation of constructs for use in cattle gene expression, or  
 CC for genetically improving cattle. The present sequence is one of the  
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The  
 CC present sequence was not shown in the specification but was obtained in  
 CC electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139



SQ Sequence 397 BP; 105 A; 84 C; 53 G; 155 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 8; Length 397;  
Best Local Similarity 83.3%; Pred. No. 5e+02; Length 397;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGCTG 24  
|||||  
Db 42 AAAAAAATAAACAATGGCGG 19  
|||||

RESULT 27  
ADX47179  
ID ADX47179 standard; cDNA; 459 BP.  
XX  
AC ADX47179;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 21919.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
PD 19-FEB-2004.  
XX  
PF 28-APR-2003; 2003US-00425114.  
XX  
PR 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX  
PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX  
DR WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
XX pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
PS Claim 1; SEQ ID NO 21919; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
XX polynucleotide consisting of a sequence encoding an amino acid sequence  
XX available in electronic form from the US patent office at  
XX ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
XX of the invention are also useful in physical arrays of molecules and as  
XX plant breeding markers. The recombinant DNA construct is useful for  
XX improving plant tolerance to cold, heat, drought, herbicides, extreme  
XX osmotic conditions, pathogens or pests, for manipulating growth rate in  
XX plant cells by modification of the cell cycle pathway, for conferring  
XX increased resistance to plant disease, for producing galactomannan,  
XX lignin or plant growth regulators, for increasing the rate of homologous  
XX recombination in plants, for improving yield by modification of  
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
XX .or by providing improved plant growth and development under at least one

CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 459 BP; 103 A; 124 C; 158 G; 74 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 459;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACATGACCAACTGGCTG 24  
|||||  
Db 1 AAAAAACAAGGACCGACTGGTCTG 24  
|||||

RESULT 28  
ADX47237  
ID ADX47237 standard; cDNA; 473 BP.  
XX  
AC ADX47237;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 21977.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
PD 19-FEB-2004.  
XX  
PF 28-APR-2003; 2003US-00425114.  
XX  
PR 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX  
PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX  
DR WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
XX pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
PS Claim 1; SEQ ID NO 21977; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
XX polynucleotide consisting of a sequence encoding an amino acid sequence  
XX available in electronic form from the US patent office at  
XX ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
XX of the invention are also useful in physical arrays of molecules and as  
XX plant breeding markers. The recombinant DNA construct is useful for  
XX improving plant tolerance to cold, heat, drought, herbicides, extreme  
XX osmotic conditions, pathogens or pests, for manipulating growth rate in  
XX plant cells by modification of the cell cycle pathway, for conferring  
XX increased resistance to plant disease, for producing galactomannan,  
XX lignin or plant growth regulators, for increasing the rate of homologous  
XX recombination in plants, for improving yield by modification of  
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
XX .or by providing improved plant growth and development under at least one

CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
XX invention.

SQ Sequence 473 BP; 105 A; 131 C; 169 G; 68 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 13; Length 473;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAACAATGACCACTGGGCTG 24  
Dbb 444 ATAAACAAGGACCGACTGGTCTG 467

RESULT 29  
AAS27499  
ID AAS27499 standard; cDNA; 555 BP.  
XX AC AAS27499;  
XX DT 07-NOV-2001 (first entry)  
XX DE cDNA encoding novel signal transduction pathway protein, Seq ID 534.  
XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
XX acquired immune deficiency syndrome.

XX OS Homo sapiens.  
XX PN WO200154733-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001312.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.





XX 05-APR-2001.  
XX 28-SEP-2000; 2000WO-US026524.  
XX 29-SEP-1999; 99US-0157137P.  
XX 03-NOV-1999; 99US-0163280P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI; 2001-235357/24.  
XX P-PSDB; AAG74146.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers.  
XX  
XX Claim 1; Page 2666; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the  
XX proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene therapy  
XX and vaccine production. N and P may be used in the prevention, diagnosis  
XX and treatment of diseases associated with inappropriate P expression. For  
XX example, N and P may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of P by expressing inactive proteins or to  
XX supplement the patient's own production of P. Additionally, N may be used  
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids  
XX into a host cell and culturing the cell to express the proteins. N and P  
XX can be used in the prevention, diagnosis and treatment of colorectal  
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent  
XX sequences used in the exemplification of the present invention. N.B.  
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at  
XX time of publication, meaning no sequences are present for SEQ ID NO:1027  
XX to 1052, 7921 and 7922  
XX  
XX Sequence 669 BP; 175 A; 185 C; 142 G; 167 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 4; Length 669;  
Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAACAATGACCAACTGGGCTGT 25  
DB 140 AAATTCATCAATGACCAACTGGGCTGT 163  
RESULT 33  
AAZ15784  
ID AAZ15784 standard; cDNA; 797 BP.  
XX  
XX AAZ15784;  
XX  
XX 12-OCT-1999 (first entry)  
XX  
XX Human gene expression product cDNA sequence SEQ ID NO:3253.  
XX  
XX Human; gene; gene expression product; diagnosis; therapy; probe;  
XX detection; mapping; tissue typing; profiling; forensic; cancer;  
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9938972-A2.  
XX  
XX 05-AUG-1999.  
XX  
XX 28-JAN-1999; 99WO-US001619.  
XX  
XX 28-JAN-1998; 98US-0072910P.  
XX

PR 24-FEB-1998; 98US-0075954P.  
PR 31-MAR-1998; 98US-0080114P.  
PR 03-APR-1998; 98US-0080515P.  
PR 03-APR-1998; 98US-0080666P.  
PR 21-OCT-1998; 98US-0105234P.  
PR 28-OCT-1998; 98US-0105877P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis WA, Garcia PD, Sudduth-Klinger J;  
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kaseam R;  
XX Lameon G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
XX Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;  
XX WPI; 1999-494092/41.  
XX  
XX Novel human genes and their expression products which are differentially  
XX expressed in different cell types.  
XX  
XX Claim 1; Page 1561; 2479pp; English.  
XX  
XX The present invention describes a library of human polynucleotides  
XX comprising the sequences given in AAZ1532 to AAZ1779. Also described is  
XX a method of detecting differentially expressed genes correlated with the  
XX cancerous state of a mammalian cell, comprising detecting at least one  
XX differentially expressed gene product in a test sample from a cell  
XX suspected of being cancerous, where the gene product is encoded by one of  
XX the 5248 polynucleotide sequences given in AAZ1532 to AAZ1779. The  
XX polynucleotides can be used as a source of primers and probes, which can  
XX be used for a variety of purpose, e.g. detection of expression levels,  
XX mapping, tissue typing or profiling, forensics, genetic analysis and  
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides  
XX can be used for raising antibodies for experimental, diagnostic and  
XX therapeutic purposes. The polynucleotides may also be used to construct  
XX arrays for diagnostics (which may be used to determine function of an  
XX encoded protein); and to detect differences in expression levels between  
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to  
XX identify a genetic predisposition or susceptibility to a disease such as  
XX cancer). The polynucleotides of the invention are especially used in the  
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,  
XX and lung cancer. The polynucleotides can also be used to screen for  
XX peptide analogues and antagonists  
XX  
XX Sequence 797 BP; 194 A; 217 C; 162 G; 203 T; 0 U; 21 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 2; Length 797;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAACAATGACCAACTGGGCTGT 25  
DB 385 AAATTCATCAATGACCAACTGGGCTGT 408  
RESULT 34  
ADJ74351/C  
ID ADJ74351 standard; cDNA; 875 BP.  
XX  
XX ADJ74351;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX Rat cDNA sequence required for viral infection SeqID 265.  
XX  
XX ss; viral infection; tumour suppression; bacterial; parasitic growth;  
XX gene trap; serum survival factor; cytostatic; virucidal; antibacterial;  
XX antiparasitic; rat.  
XX  
XX Rattus sp.  
XX  
XX WO2004010925-A2.  
XX

```
PD 05-FEB-2004.
XX
XX
PF 02-MAY-2003; 2003WO-US013743.
XX
XX
PR 02-MAY-2002; 2002US-0377136P.
XX
XX
PA (UYVA-) UNIV VANDERBILT.
XX
XX
PI Rubin DH;
XX
XX
DR WPI; 2004-143721/14.
XX
XX
PT Novel gene useful for preparing a composition for reducing or inhibiting
PT viral infection or for suppressing a malignant phenotype in a cell.
XX
XX
PS Claim 1; SEQ ID NO 265; 662pp; English.
XX
XX
CC This invention relates to novel mammalian genes involved in viral
CC infection and tumour suppression. Specifically, it refers to methods for
CC identifying cellular genes that are required for viral, bacterial or
CC parasitic growth, as well as genes used for tumour progression, but that
CC are not essential for cell survival. The present invention describes a
CC 'gene trap' method and screening/selection process that can isolate
CC those polynucleotides that are associated with a specific process of
CC interest. Furthermore, this method relies on the core discovery that
CC virally infected cells become dependent on a serum survival factor that
CC does not affect non-infected cells. Accordingly, identification of these
CC factors and the appropriate inhibitors thereof, provides a means to treat
CC and/or prevent viral, bacterial or parasitic growth and infection, as
CC well as tumour growth. The compositions of this invention exhibit
CC cytostatic, virucidal, antibacterial and antiparasitic activities. This
CC polynucleotide is a rat CDNA sequence that is necessary for viral
CC infection, given in an exemplification of the invention.
XX
XX
SQ Sequence 875 BP; 221 A; 210 C; 205 G; 211 T; 0 U; 28 Other;
Query Match 70.4%; Score 17.6; DB 12; Length 875;
Best Local Similarity 83.3%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACATGACCAACTGGGCTG 24
Db 88 AAAAAAATGACACATTGGGCTG 65
RESULT 35
ADX47961/c
ID ADX47961 standard; cDNA; 1163 BP.
AC ADX47961;
XX
XX
DT 21-APR-2005 (first entry)
XX
XX
DE Plant full length insert polynucleotide seqid 22701.
XX
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX
XX
OS Unidentified.
XX
XX
PN US2004034888-A1.
XX
XX
PD 19-FEB-2004.
XX
XX
PF 28-APR-2003; 2003US-00425114.
XX
XX
PR 06-MAY-1999; 99US-00304517.
XX
XX
PR 05-NOV-2001; 2001US-00985678.
XX
XX
PA (LIUJ/) LIU J.
PA (ZHOUJ/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABAJ/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
XX
DR WPI; 2004-180133/17.
XX
XX
PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX
PS Claim 1; SEQ ID NO 22701; 15pp; English.
XX
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX
SQ Sequence 1163 BP; 249 A; 318 C; 336 G; 260 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 13; Length 1163;
Best Local Similarity 83.3%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACATGACCAACTGGGCTG 24
Db 57 ATAAACCAAGGACCGACTGGTCTG 34
RESULT 36
AAD41402
ID AAD41402 standard; cDNA; 1670 BP.
AC AAD41402;
XX
XX
DT 30-OCT-2002 (first entry)
XX
XX
DE Human NZMS-15 cDNA.
XX
XX
KW Human; enzyme; NZMS-15; cell proliferative disorder; hepatitis; cancer;
KW psoriasis; leukaemia; autoimmune disorder; diabetes; arteriosclerosis;
KW acquired immune deficiency syndrome; cardiovascular disorder; anorexia;
KW asthma; hypertension; gastrointestinal disorder; reproductive disorder;
KW gastritis; neurological disorder; epilepsy; pulmonary disorder; AIDS;
KW dementia; embolism; gene therapy; eye disorder; transgenic; gene; ss.
XX
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1512
FT /*tag= a
FT /product= "NZMS-15 protein"
```

```
FT sig_peptide 1. .99
FT PT /*tag= b
FT mat_peptide 100. .1509
FT FT /*tag= c
FT FT /product= "Mature NZMS-15 protein"
XX
XX WO200246385-A2.
XX
XX 13-JUN-2002.
XX
XX 04-DEC-2001; 2001WO-US047432.
XX
XX 07-DEC-2000; 2000US-0251824P.
XX PR 08-DEC-2000; 2000US-0254312P.
XX PR 14-DEC-2000; 2000US-0255773P.
XX PR 15-DEC-2000; 2000US-0255940P.
XX PR 15-DEC-2000; 2000US-0256188P.
XX PR 21-DEC-2000; 2000US-0257488P.
XX PR 19-JAN-2001; 2001US-0262839P.
XX PR 26-JAN-2001; 2001US-0264402P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Tang YT, Griffin JA, Yue H, Lee EA, Baughn MR, Duggan BM;
PI Walia NK, Lee S, Ramkumar J, Warren BA, Gandhi AR, Lu DM, Lu Y;
PI Yao MG, Ding L, Tribouley CM, Sanjanwala MM, Arvizu C, Hillman JL;
XX
XX WPI; 2002-537565/57.
XX P-PSDB; AAE25391.
XX
XX Novel human enzyme, NZMS useful in diagnosis, prevention or treatment of
PT cell proliferative, autoimmune/inflammatory, cardiovascular,
PT gastrointestinal, neurological, pulmonary, reproductive and eye
PT disorders.
XX
XX Claim 88; Page 170-171; 173pp; English.
XX
XX The invention relates to human enzymes designated NZMS and nucleic acid
CC molecules encoding such proteins. Sequences of the invention are useful
CC for diagnosing, treating or preventing disorders associated with aberrant
CC expression of NZMS. The disorders treated include cell proliferative
CC disorders such as hepatitis, psoriasis, cancer (e.g. leukaemia),
CC autoimmune disorders such as diabetes, acquired immune deficiency
CC syndrome (AIDS), cardiovascular disorders such as arteriosclerosis,
CC hypertension), gastrointestinal disorders (e.g. anorexia, gastritis),
CC neurological disorders (e.g. epilepsy, dementia), pulmonary disorders
CC (e.g. embolism, asthma), reproductive or eye disorders. Polypeptides of
CC the invention is useful in a number of drug screening techniques and to
CC analyse the proteome of a tissue or cell type. They are also useful as
CC elements on a microarray. Polynucleotides of the invention are useful for
CC creating knockin humanised animals or transgenic animals to model human
CC diseases, in somatic or germline gene therapy, to generate a transcript
CC image of a tissue or cell type, for detecting difference in the
CC chromosomal location due to translocation or inversion among normal,
CC carrier or affected individuals and as hybridisation probes for mapping
CC naturally occurring genomic sequences. The present sequence is human NZMS
CC cDNA
XX
XX Sequence 1670 BP; 460 A; 355 C; 438 G; 417 T; 0 U; 0 Other;
SQ
XX
XX Query Match 70.4%; Score 17.6; DB 6; Length 1670;
XX Best Local Similarity 83.3%; Pred. No. 6.2e+02;
XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX Qy 2 AAAAACCAATGATCCAGGGCTGT 25
XX |||||||
XX Db 1032 AAAAACCAATGATCCAGGGCTGT 1055
XX
XX RESULT 37
XX AAS27126
XX ID AAS27126 standard; cDNA; 1921 BP.
XX
```

AC AAS27126;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX cDNA encoding novel signal transduction pathway protein, Seq ID 161.  
XX  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
XX acquired immune deficiency syndrome.  
XX  
XX Homo sapiens.  
XX  
XX WO200154733-A1.  
XX PN  
XX 02-AUG-2001.  
XX PD  
XX  
XX 17-JAN-2001; 2001WO-US001312.  
XX PF  
XX 31-JAN-2000; 2000US-0179065P.  
XX PR 04-FEB-2000; 2000US-0180628P.  
XX PR 24-FEB-2000; 2000US-0184664P.  
XX PR 02-MAR-2000; 2000US-0186350P.  
XX PR 16-MAR-2000; 2000US-0189874P.  
XX PR 17-MAR-2000; 2000US-0190076P.  
XX PR 18-APR-2000; 2000US-0198123P.  
XX PR 19-MAY-2000; 2000US-0205515P.  
XX PR 07-JUN-2000; 2000US-0209467P.  
XX PR 28-JUN-2000; 2000US-0214886P.  
XX PR 30-JUN-2000; 2000US-0215135P.  
XX PR 07-JUL-2000; 2000US-0216647P.  
XX PR 07-JUL-2000; 2000US-0216880P.  
XX PR 11-JUL-2000; 2000US-0217487P.  
XX PR 11-JUL-2000; 2000US-0217496P.  
XX PR 14-JUL-2000; 2000US-0218290P.  
XX PR 26-JUL-2000; 2000US-0220963P.  
XX PR 26-JUL-2000; 2000US-0220964P.  
XX PR 14-AUG-2000; 2000US-0224518P.  
XX PR 14-AUG-2000; 2000US-0224519P.  
XX PR 14-AUG-2000; 2000US-0225213P.  
XX PR 14-AUG-2000; 2000US-0225214P.  
XX PR 14-AUG-2000; 2000US-0225266P.  
XX PR 14-AUG-2000; 2000US-0225267P.  
XX PR 14-AUG-2000; 2000US-0225268P.  
XX PR 14-AUG-2000; 2000US-0225270P.  
XX PR 14-AUG-2000; 2000US-0225447P.  
XX PR 14-AUG-2000; 2000US-0225757P.  
XX PR 14-AUG-2000; 2000US-0225758P.  
XX PR 14-AUG-2000; 2000US-0225759P.  
XX PR 18-AUG-2000; 2000US-0226279P.  
XX PR 22-AUG-2000; 2000US-0226681P.  
XX PR 22-AUG-2000; 2000US-0226988P.  
XX PR 22-AUG-2000; 2000US-0227182P.  
XX PR 23-AUG-2000; 2000US-0227009P.  
XX PR 30-AUG-2000; 2000US-0228924P.  
XX PR 01-SEP-2000; 2000US-0229287P.  
XX PR 01-SEP-2000; 2000US-0229343P.  
XX PR 01-SEP-2000; 2000US-0229344P.  
XX PR 01-SEP-2000; 2000US-0229345P.  
XX PR 05-SEP-2000; 2000US-0229509P.  
XX PR 05-SEP-2000; 2000US-0229513P.  
XX PR 06-SEP-2000; 2000US-0230437P.  
XX PR 06-SEP-2000; 2000US-0230438P.  
XX PR 08-SEP-2000; 2000US-0231242P.  
XX PR 08-SEP-2000; 2000US-0231243P.  
XX PR 08-SEP-2000; 2000US-0231244P.  
XX PR 08-SEP-2000; 2000US-0231413P.





KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;  
KW immunosuppressive agent; adjuvant; enhance immune response;  
KW higher affinity antibody induction;  
KW increased serum immunoglobulin concentration.  
XX

OS Homo sapiens.

PN US2002168711-A1.

XX 14-NOV-2002.

XX 17-JAN-2001; 2001US-00764868.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 28-JUN-2000; 2000US-0214886P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 22-AUG-2000; 2000US-0226868P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 08-SEP-2000; 2000US-0229513P.

XX 21-SEP-2000; 2000US-0234223P.

XX 21-SEP-2000; 2000US-0234274P.

XX 25-SEP-2000; 2000US-0234997P.

XX 27-SEP-2000; 2000US-0235834P.

XX 29-SEP-2000; 2000US-0236327P.

XX 29-SEP-2000; 2000US-0236367P.

XX 29-SEP-2000; 2000US-0236368P.

XX 29-SEP-2000; 2000US-0236369P.

XX 29-SEP-2000; 2000US-0236370P.

XX 02-OCT-2000; 2000US-0236802P.

XX 02-OCT-2000; 2000US-0237037P.

XX 02-OCT-2000; 2000US-0237038P.

XX 02-OCT-2000; 2000US-0237039P.

XX 02-OCT-2000; 2000US-0237040P.

XX 13-OCT-2000; 2000US-0239935P.

XX 20-OCT-2000; 2000US-0240960P.

XX 20-OCT-2000; 2000US-0241785P.

XX 20-OCT-2000; 2000US-0241809P.

XX 01-NOV-2000; 2000US-0244617P.

XX 17-NOV-2000; 2000US-0249299P.

XX 08-DEC-2000; 2000US-0251856P.

XX 08-DEC-2000; 2000US-0251868P.

XX 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.

XX (RUBE/) RUBEN S M.

XX (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-719985/68.

XX P-PSDB; ADB93917.

PT New isolated polypeptide useful for diagnosing and treating  
PT immunosuppressive conditions such as autoimmune disease and Parkinson's  
PT disease.  
XX

XX Claim 3; SEQ ID NO 161; 345pp; English.

XX The invention relates to an isolated polypeptide. The polypeptide is  
CC useful for diagnosing a pathological condition or a susceptibility to a  
CC pathological condition in a subject, by determining the presence or  
CC amount of expression of the polypeptide in a biological sample and  
CC diagnosing a pathological condition or a susceptibility to a pathological  
CC condition based on the presence or amount of expression of the  
CC polypeptide. The polypeptide is also useful for identifying a binding  
CC partner to the polypeptide, which involves contacting the polypeptide  
CC with a binding partner and determining whether the binding partner  
CC effects an activity of the polypeptide. The polypeptide or the nucleic  
CC acid encoding the polypeptide is useful for preventing, treating, or  
CC ameliorating a medical condition, which involves administering the  
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid  
CC is useful for diagnosing a pathological condition or a susceptibility to  
CC a pathological condition in a subject, which involves determining the  
CC presence or absence of a mutation in the nucleic acid, and diagnosing a  
CC pathological condition or susceptibility to a pathological condition  
CC based on the presence or absence of the mutation. The polypeptide, the  
CC nucleic acid and an antibody to the polypeptide are useful for treating  
CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal  
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,  
CC the nucleic acid and the antibody are useful as immunosuppressive agents,  
CC as adjuvants to enhance immune responses, and as agents to induce higher  
CC affinity antibodies and increase serum immunoglobulin concentrations. The  
CC present sequence represents cDNA encoding a novel human protein. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format direct from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20020168711.

XX SQ Sequence 1921 BP; 598 A; 356 C; 454 G; 513 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 10; Length 1921;

Best Local Similarity 83.3%; Pred. No. 6.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCACTGGGCTGT 25

Db 415 AAAACAATGATCGCCAGGCTGT 438

RESULT 39

AAH79150

ID AAH79150 standard; cDNA; 2229 BP.

XX AC AAH79150;

XX DT 19-NOV-2001 (first entry)

XX DE Human RNA helicase 9 encoding cDNA.

XX Human; RNA helicase 9; cytostatic; virucidal; immunomodulatory;  
KW antiinflammatory; haemostatic; malignant neoplasm; HIV; infection;  
KW human immunodeficiency virus; immunological disease; cancer;  
XX nervous system disease; ss.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 800..1042

FT /\*tag= a

FT /product= "RNA helicase 9"

FT /note= "claimed in claim 6"

XX WO200166592-A1.

XX PN 13-SEP-2001.

XX PD

XX

```
PF 26-FEB-2001; 2001WO-CN000217.
XX
PR 07-MAR-2000; 2000CN-00111929.
XX
XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
XX MAO Y, XIE Y;
PI
XX WPI; 2001-589932/66.
DR P-PSDB; AAG78128.
XX
XX Human RNA helicase 9 and encoded polynucleotide, for diagnosis and
PT treatment of e.g. nervous system diseases, malignant neoplasm, hemopathy,
PT HIV infection, immunological diseases and various inflammations.
XX
XX Claim 6; Page 29-30; 34pp; Chinese.
XX
XX The invention relates to the human RNA helicase 9 and the encoding
CC polynucleotide with cytostatic, virucidal, immunomodulatory,
CC antiinflammatory and haemostatic activity. The polypeptide and encoded
CC polynucleotide are applicable in diagnosis and treatment of malignant
CC neoplasm, haemopathy, HIV infection, immunological diseases, various
CC inflammations, cancer and nervous system diseases
XX
XX Sequence 2229 BP; 666 A; 459 C; 477 G; 627 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 4; Length 2229;
Best Local Similarity 83.3%; Pred. No. 6.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACATGACCACTGGGCTG 24
DB 1935 AAAAAAAGGTACAACTGGGCTG 1958
RESULT 40
AD035596
ID AD035596 standard; DNA; 2229 BP.
XX
AC AD035596;
XX
DT 26-AUG-2004 (first entry)
XX
DE Novel mouse gene sequence #269.
XX
XX mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;
KW ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;
KW viral disorder; ds; gene.
XX
XX Mus sp.
OS
XX WO2004046310-A2.
XX
XX 03-JUN-2004.
XX
XX 24-OCT-2003; 2003WO-US033948.
XX
XX 15-NOV-2002; 2002US-0426916P.
PR
XX 04-DEC-2002; 2002US-0431158P.
PR
XX 05-DEC-2002; 2002US-0431445P.
PR
XX 05-DEC-2002; 2003US-0431606P.
PR
XX 09-JUN-2003; 2003US-0476621P.
PR
XX 08-JUL-2003; 2003US-0485217P.
PR
XX 08-AUG-2003; 2003US-0485359P.
PR
XX 08-AUG-2003; 2003US-0493332P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX Williams LT, Chu K, Lee E, Hestir K, Hayashizaki Y, Kamiya M;
PI
XX WPI; 2004-431966/40.
XX
XX New mouse nucleic acid molecules and polypeptides, useful for treating
PT cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart
PT disease or thrombosis.
XX
XX Claim 1; SEQ ID NO 269; 263pp; English.
XX
XX The invention comprises 744 novel mouse DNA sequences (genes). The DNA
CC sequences of the invention are useful for treating cancer, psoriasis,
CC ulcerative colitis, inflammation, ischaemic heart disease, thrombosis,
CC immune disorders, bacterial disorders and viral disorders. The present
CC nucleic acid represents a mouse DNA sequence of the invention. NOTE: The
CC present DNA sequence is not shown in the specification, but has been
XX retrieved from the WIPO website.
XX
XX Sequence 2229 BP; 659 A; 418 C; 485 G; 667 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 12; Length 2229;
Best Local Similarity 83.3%; Pred. No. 6.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACATGACCACTGGGCTG 24
DB 341 AAAAAATCAATGCCAACTGGGCAG 364
RESULT 41
ABA20469/c
ID ABA20469 standard; DNA; 2449 BP.
XX
AC ABA20469;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12800.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
OS
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
XX 17-MAR-2000; 2000US-0190076P.
PR
XX 18-APR-2000; 2000US-0198123P.
PR
XX 19-MAY-2000; 2000US-0205515P.
PR
XX 07-JUN-2000; 2000US-0209467P.
PR
XX 28-JUN-2000; 2000US-0214886P.
PR
XX 30-JUN-2000; 2000US-0215135P.
PR
XX 07-JUL-2000; 2000US-0216647P.
PR
XX 07-JUL-2000; 2000US-0216880P.
PR
XX 11-JUL-2000; 2000US-0217487P.
PR
XX 11-JUL-2000; 2000US-0217496P.
PR
XX 14-JUL-2000; 2000US-0218290P.
PR
XX 26-JUL-2000; 2000US-0220963P.
PR
XX 26-JUL-2000; 2000US-0220964P.
PR
XX 14-AUG-2000; 2000US-0224518P.
PR
XX 14-AUG-2000; 2000US-0224519P.
PR
XX 14-AUG-2000; 2000US-0225213P.
```

PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225575P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246611P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246613P.
PR	18-AUG-2000;	2000US-0226279P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0226868P.	PR	17-NOV-2000;	2000US-0249209P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249210P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249211P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229334P.	PR	17-NOV-2000;	2000US-0249215P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249217P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249244P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249245P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232080P.	PR	01-DEC-2000;	2000US-0250391P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0251160P.
PR	12-SEP-2000;	2000US-02321968P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232399P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234221P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-02559678P.
PR	25-SEP-2000;	2000US-0234997P.	XX		
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0234998P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235835P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2			

```
SQ Sequence 2449 BP; 541 A; 596 C; 600 G; 712 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 5; Length 2449;
Best Local Similarity 83.3%; Pred. No. 6.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
Db 1696 AAGAACAAATGACCAAAATGACTG 1673

RESULT 42
AAL57079
ID AAL57079 standard; cDNA; 2532 BP.
XX
AC AAL57079;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human sperm generation associated factor 86-90 cDNA sequence.
XX
KW Human sperm generation associated factor 86-90; osteoma; leukaemia; gene;
XX ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2679
FT /*tag= a
FT /product= "transporter protein"
XX
FN WO2003076645-A2.
XX
PD 18-SEP-2003.
XX
PF 05-MAR-2003; 2003WO-US006670.
XX
PR 05-MAR-2002; 2002US-0361343P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Yan X, Yan C, Neelam B;
XX
WPI; 2003-748395/70.
DR P-PSDB; ADE11167.
XX
PT New human transporter proteins and nucleic acid molecules, useful for
PT diagnosing, preventing or treating disorders associated with aberrant
PT expression of the transporter proteins, e.g. tumors, or in
PT pharmacogenomic analysis.
XX
PS Claim 4; SEQ ID NO 1; 330pp; English.
XX
CC The invention relates to a novel isolated protein comprising a fully
CC defined sequence of 892 amino acids given in the specification, an
CC allelic variant or an orthologue of the protein and/or a fragment of the
CC protein. A protein of the invention has cytostatic activity, and may have
CC a use in gene therapy. The protein or nucleic acid molecule encoding it
CC is useful as a model for the development of human therapeutic targets,
CC aids in the identification of therapeutic proteins, serve as targets for
CC the development of human therapeutic agents that modulate transporter
CC activity in cells and tissues that express the transporter, or in
CC pharmacogenomic analysis. The protein may be used in substantial and
CC specific assays, in raising antibodies or in eliciting another immune
CC response, as a reagent in assays designed to quantitatively determine the
CC levels of the protein in biological fluids, and as markers for tissues in
CC which the protein is expressed. The nucleic acid molecules may be used as
CC probes, primers, chemical intermediates and in biological assays. These
CC may also be used in diagnosing, preventing or treating disorders
CC associated with human transporter protein, such as tumours. The present
CC sequence encodes the human transporter protein of the invention.
XX
SQ Sequence 2679 BP; 800 A; 517 C; 673 G; 689 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 10; Length 2679;
Best Local Similarity 83.3%; Pred. No. 6.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCTGT 25
Db 1194 AAAAAACAATGATCGCCAGGGCTGT 1217

RESULT 44
ABT16031
ID ABT16031 standard; DNA; 2789 BP.
XX
AC ABT16031;
XX
```

DT 28-MAR-2003 (first entry)

XX NOVX related polynucleotide SEQ ID No 29.

DE

XX Antidiabetic; anorectic; virucide; antibacterial; fungicide; nootropic; nootropic; neuroprotective; antiparkinsonian; antilipidemic; antilipidemic; NOVX-associated disorder; metabolic disorder; diabetes; anorexia; obesity; infectious disease; cancer-associated cachexia; immune disease; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; haematopoietic disorder; cancer; dyslipidaemia; metabolic disturbance; neurogenesis; cell differentiation; cell proliferation; haematopoiesis; wound healing; angiogenesis; gene therapy; chromosome mapping; tissue typing; preventive medicine; pharmacogenomic; NOVX; gene; ds.

XX Unidentified.

OS

XX WO200299062-A2.

PN

XX 12-DEC-2002.

PD

XX 04-JUN-2002; 2002WO-US017559.

XX

XX 04-JUN-2001; 2001US-0295607P.

XX

PR 06-JUN-2001; 2001US-0296404P.

PR

PR 06-JUN-2001; 2001US-0296418P.

PR

PR 07-JUN-2001; 2001US-0296575P.

PR

PR 11-JUN-2001; 2001US-0297414P.

PR

PR 12-JUN-2001; 2001US-0297567P.

PR

PR 12-JUN-2001; 2001US-0297573P.

PR

PR 14-JUN-2001; 2001US-0298285P.

PR

PR 15-JUN-2001; 2001US-0298528P.

PR

PR 15-JUN-2001; 2001US-0298556P.

PR

PR 18-JUN-2001; 2001US-0299133P.

PR

PR 19-JUN-2001; 2001US-0299230P.

PR

PR 21-JUN-2001; 2001US-0299949P.

PR

PR 22-JUN-2001; 2001US-0300177P.

PR

PR 28-JUN-2001; 2001US-0301530P.

PR

PR 28-JUN-2001; 2001US-0301550P.

PR

PR 03-JUL-2001; 2001US-0302951P.

PR

PR 12-SEP-2001; 2001US-0318711P.

PR

PR 25-SEP-2001; 2001US-0324687P.

PR

PR 24-OCT-2001; 2001US-0339266P.

PR

PR 16-NOV-2001; 2001US-0337524P.

PR

PR 14-DEC-2001; 2001US-0341143P.

PR

PR 21-FEB-2002; 2002US-0358643P.

PR

PR 21-FEB-2002; 2002US-0359151P.

PR

PR 28-FEB-2002; 2002US-0361195P.

PR

PR 05-MAR-2002; 2002US-0361964P.

PR

PR 10-APR-2002; 2002US-0371346P.

PR

PR 10-APR-2002; 2002US-0371523P.

PR

PR 03-JUN-2002; 2002US-00161493.

XX

XX (CURA-) CURAGEN CORP.

XX

XX Anderson DW, Zerhusen BD, Li L, Zhong M, Casman SJ, Gerlach VL; Shimkets RA, Gorman L, Pena CE, Kekuda R, Patturajan M, Spytek KA; Leite MW, Rastelli L, Macdougall JR, Taupier RJ, Guo X, Miller CE; Shenoy SG, Hjalt T, Voss EZ, Boldog FL, Malyankar UM, Padigar M; Ji W, Smithson G, Edinger SR, Millet I, Ellerman K;

XX

XX WPI; 2003-140607/13.

DR

DR P-PSDB; ABJ19332.

XX

XX New isolated NOVX polypeptides and polynucleotides, useful for preventing, diagnosing or treating NOVX-associated disorders, e.g. obesity, cancer, Parkinson's disease, infections, immune disorders, or various dyslipidaemias.

XX

XX Example 12; Page 122; 461pp; English.

XX

XX The invention relates to an isolated polypeptide comprising any of the 36 86-1370 residue amino acid sequences, given in the specification, a mature form of them, or a sequence that is at least 95 % identical to, or

CC

CC having one or more conservative amino acid substitutions in one of the 36 amino acid sequences. The polypeptides, nucleic acid molecules and antibodies of the invention are useful in the manufacture of a medicament for treating a syndrome associated with a human disease, preferably a NOVX-associated disorder. The nucleic acid molecules, polypeptides and antibodies are useful for treating, preventing or diagnosing diseases such as metabolic disorders, diabetes, obesity, infectious diseases (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer, Parkinson's disease, neurodegenerative disorders, Alzheimer's disease, cancer and various dyslipidaemias, or metabolic disturbances associated with obesity, metabolic X syndrome, and wasting disorders. The nucleic acids and polypeptides may also be used as targets for the identification of small molecules that modulate or inhibit e.g. neurogenesis, cell differentiation, cell proliferation, haematopoiesis, wound healing and angiogenesis, in gene therapy, in generation of antibodies that bind immunospecifically to NOVX substances for use in therapeutic or diagnostic methods. The nucleic acids are further used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine, and pharmacogenomics. This polynucleotide represents a NOVX related DNA sequence of the invention

XX

SQ Sequence 2789 BP; 830 A; 538 C; 695 G; 726 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 8; Length 2789;

Best Local Similarity 83.3%; Pred. No. 6.8e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACACATGACCAACTGGCTGT 25

DB 1257 AAAACACATGATCGCCAGGGCTGT 1280

RESULT 45

ADO41663

ID ADO41663 standard; cDNA; 2789 BP.

XX

XX ADO41663;

XX

XX 15-JUL-2004 (first entry)

XX

XX Novel human polypeptide NOV12b cDNA.

XX

XX cardiant; antiarteriosclerotic; hypotensive; immunosuppressive; dermatological; anorectic; cytostatic; antidiabetic; haemostatic; anti-HIV; antiasthmatic; antibacterial; virucide; neuroprotective; nootropic; antiparkinsonian; antilipemic; analgesic; antianginal; gene therapy; vaccine; NOVX-associated disorder; cardiomyopathy; atherosclerosis; hypertension; scleroderma; obesity; cancer; diabetes; haemophilia; graft-versus-host disease; AIDS; asthma; Crohn's disease; multiple sclerosis; infection; anorexia; cancer-associated cachexia; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; haematopoietic disorder; dyslipidemia; wasting disorder; chromosome mapping; tissue typing; preventive medicine; pharmacogenomic; vaccine; human; gene; ss.

XX

XX Homo sapiens.

OS

XX US2004018555-A1.

PN

XX 29-JAN-2004.

XX

XX 03-JUN-2002; 2002US-00161493.

XX

XX 04-JUN-2001; 2001US-0295607P.

PR

PR 06-JUN-2001; 2001US-0296404P.

PR

PR 06-JUN-2001; 2001US-0296418P.

PR

PR 07-JUN-2001; 2001US-0296575P.

PR

PR 11-JUN-2001; 2001US-0297414P.

PR

PR 12-JUN-2001; 2001US-0297567P.

PR

PR 12-JUN-2001; 2001US-0297573P.

PR

PR 14-JUN-2001; 2001US-0298285P.

PR

PR 15-JUN-2001; 2001US-0298528P.

PR

PR 15-JUN-2001; 2001US-0298556P.  
 PR 18-JUN-2001; 2001US-0298133P.  
 PR 19-JUN-2001; 2001US-0298230P.  
 PR 21-JUN-2001; 2001US-0299949P.  
 PR 22-JUN-2001; 2001US-0300177P.  
 PR 28-JUN-2001; 2001US-0301530P.  
 PR 28-JUN-2001; 2001US-0301550P.  
 PR 03-JUL-2001; 2001US-0302951P.  
 PR 12-SEP-2001; 2001US-0318771P.  
 PR 28-SEP-2001; 2001US-0325687P.  
 PR 24-OCT-2001; 2001US-0332666P.  
 PR 16-NOV-2001; 2001US-0337524P.  
 PR 14-DEC-2001; 2001US-0341143P.  
 PR 21-FEB-2002; 2002US-0358643P.  
 PR 21-FEB-2002; 2002US-0359151P.  
 PR 28-FEB-2002; 2002US-0361195P.  
 PR 05-MAR-2002; 2002US-0361964P.  
 PR 10-APR-2002; 2002US-0371346P.  
 PR 10-APR-2002; 2002US-0371523P.  
 XX  
 PA (ANDE/) ANDERSON D W.  
 PA (ZERH/) ZERHUSEN B D.  
 PA (LILL/) LI L.  
 PA (ZHON/) ZHONG M.  
 PA (CASH/) CASMAN S J.  
 PA (GERL/) GERLACH V.  
 PA (SHIM/) SHIMKETS R A.  
 PA (GORM/) GORMAN L.  
 PA (PENA/) PENA C E A.  
 PA (KEKU/) KEKUDA R.  
 PA (PAT/) PATURAJAN M.  
 PA (SPYT/) SPYTEK K A.  
 PA (LEIT/) LEITE M W.  
 PA (RAST/) RASTELLI L.  
 PA (MACD/) MACDOUGALL J R.  
 PA (TAUP/) TAUPIER R J.  
 PA (GUOX/) GUO X S.  
 PA (MILL/) MILLER C E.  
 PA (SHEN/) SHENOY S G.  
 PA (HJAL/) HJALT T.  
 PA (VOSS/) VOSS E Z.  
 PA (BOLD/) BOLDOG F L.  
 PA (MALY/) MALYANKAR U M.  
 PA (PADI/) PADIGARU M.  
 PA (JIWW/) JI W.  
 PA (SMIT/) SMITHSON G.  
 PA (EDIN/) EDINGER S R.  
 PA (MILL/) MILLET I.  
 PA (ELLE/) ELLERMAN K.  
 XX  
 PI Anderson DW, Zerhusen BD, Li L, Zhong M, Casman SJ, Gerlach V;  
 PI Shimkets RA, Gorman L, Pena CEA, Kekuda R, Patturajan M, Spytek KA;  
 PI Leite MW, Rastelli L, MacDougall JR, Taupier RJ, Guo XS, Miller CB;  
 PI Shenoy SG, Hjalt T, Voss EZ, Boldog FL, Malvankar UM, Padigar M;  
 PI Ji W, Smithson G, Edinger SR, Millet I, Ellerman K;  
 XX  
 WP1; 2004-122030/12.  
 DR P-PSDB; ADO41664.  
 DR  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, tissue typing  
 PT atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing  
 PT or pharmacogenomics.  
 XX  
 PS Claim 20; SEQ ID NO 29; 322pp; English.  
 XX  
 CC The invention describes a new polypeptide comprising: any of the 36 fully  
 CC defined amino acid sequences (e.g. 860, 919 or 681 amino acids) given in  
 CC the specification; a mature form of (a); a sequence that is at least 95%  
 CC identical to (a); or a sequence comprising one or more conservative  
 CC substitutions in (a). The polypeptide is useful in the manufacture of a  
 CC medicament for treating a syndrome associated with a human disease, the  
 CC disease selected from a pathology associated with the polypeptide. These

CC may also be used in diagnosing, treating or preventing NOVX-associated  
 CC disorders such as cardiomyopathy, atherosclerosis, hypertension,  
 CC scleroderma, obesity, cancer, diabetes, haemophilia, graft-versus-host  
 CC disease, AIDS, asthma, Crohn's disease, multiple sclerosis, infections,  
 CC anorexia, cancer-associated cachexia, neurodegenerative disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease), haematopoietic disorders  
 CC dyslipidemias and other wasting disorders associated with chronic  
 CC diseases. The nucleic acids are also used as hybridisation probes, in  
 CC chromosome mapping, tissue typing, preventive medicine, and  
 CC pharmacogenomics. The polypeptides are also useful as vaccines. This  
 CC sequence encodes a novel human polypeptide of the invention.  
 XX  
 SQ Sequence 2789 BP; 830 A; 538 C; 695 G; 726 T; 0 U; 0 Other;  
  
 Query Match 70.4%; Score 17.6; DB 12; Length 2789;  
 Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
 QY 2 AAAAACAATGACCAACTGGGCTGT 25  
 |||||  
 Db 1257 AAAAACAATGATGCCAGGGCTGT 1280  
  
 RESULT 46  
 ABT16030  
 ID ABT16030 standard; DNA; 2812 BP.  
 XX  
 AC ABT16030;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE NOVX related polynucleotide SEQ ID NO 27.  
 XX  
 KW Antidiabetic; anorectic; virucide; antibacterial; fungicide; nootropic;  
 KW protozoacide; neuroprotective; antiparkinsonian; antileptemic;  
 KW NOVX-associated disorder; metabolic disorder; diabetes; anorexia;  
 KW obesity; infectious disease; cancer-associated cachexia; immune disorder;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW haematopoietic disorder; cancer; dyslipidaemia; metabolic disturbance;  
 KW neurogenesis; cell differentiation; cell proliferation; haematopoiesis;  
 KW wound healing; angiogenesis; gene therapy; chromosome mapping;  
 KW tissue typing; preventive medicine; pharmacogenomic; NOVX; gene; ds.  
 XX  
 OS Unidentified.  
 XX  
 XX WO200299062-A2.  
 XX  
 PD 12-DEC-2002.  
 XX  
 XX 04-JUN-2002; 2002WO-US017559.  
 XX  
 XX 04-JUN-2001; 2001US-0295607P.  
 PR 06-JUN-2001; 2001US-0296404P.  
 PR 06-JUN-2001; 2001US-0296418P.  
 PR 07-JUN-2001; 2001US-0296575P.  
 PR 11-JUN-2001; 2001US-0297414P.  
 PR 12-JUN-2001; 2001US-0297567P.  
 PR 12-JUN-2001; 2001US-0297573P.  
 PR 14-JUN-2001; 2001US-0298285P.  
 PR 15-JUN-2001; 2001US-0298528P.  
 PR 15-JUN-2001; 2001US-0298556P.  
 PR 18-JUN-2001; 2001US-0299133P.  
 PR 19-JUN-2001; 2001US-0299230P.  
 PR 21-JUN-2001; 2001US-0299949P.  
 PR 22-JUN-2001; 2001US-0300177P.  
 PR 28-JUN-2001; 2001US-0301530P.  
 PR 28-JUN-2001; 2001US-0301550P.  
 PR 03-JUL-2001; 2001US-0302951P.  
 PR 12-SEP-2001; 2001US-0318771P.  
 PR 25-SEP-2001; 2001US-0324687P.  
 PR 24-OCT-2001; 2001US-0332666P.  
 PR 16-NOV-2001; 2001US-0337524P.  
 PR 14-DEC-2001; 2001US-0341143P.  
 PR







```

XX OS Homo sapiens.
XX FH Key
XX FT CDS
XX FT Location/Qualifiers
XX FT 252..2057
XX FT /*tag= a
XX FT /product= "ARP30"
XX PN WO2003060148-A2.
XX PD 24-JUL-2003.
XX PF 15-JAN-2003; 2003WO-US001457.
XX PR 15-JAN-2002; 2002US-00053248.
XX PA (SYST-) INST SYSTEMS BIOLOGY.
XX PI Lin B;
XX XX WPI; 2003-587287/55.
XX DR P-PSDB; ABR82449.
XX PT Diagnosing or predicting susceptibility to a prostate neoplastic
XX PT condition by contacting a specimen from the individual with an ARP15
XX PT binding agent that selectively binds an ARP15 polypeptide.
XX PS Claim 134; Page 210-214; 227pp; English.
XX CC The invention relates to diagnosing or predicting susceptibility to a
XX CC prostate neoplastic condition. The method involves (a) contacting a
XX CC specimen from the individual with an androgen responsive prostate
XX CC specific (ARP)15 binding agent that selectively binds an ARP15
XX CC polypeptide; (b) determining a test expression level of ARP15 polypeptide
XX CC in the specimen; and (c) comparing the test expression level to a non-
XX CC neoplastic control expression level of ARP15 polypeptide, where an
XX CC altered test expression level as compared to the control expression level
XX CC indicates the presence of a prostate neoplastic condition in the
XX CC individual. The method is useful for diagnosing or predicting
XX CC susceptibility to a prostate neoplastic condition or for treating or
XX CC reducing severity of a prostate neoplastic condition. The present
XX CC sequence represents a human ARP30 polypeptide encoding cDNA
XX SQ Sequence 3318 BP; 842 A; 987 C; 828 G; 661 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 9; Length 3318;
Best Local Similarity 83.3%; Pred. No. 6.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACCAATGACCAACTGGGCTGT 25
Db 2802 AAATTCAATCACCACCACTGTGCTGT 2825

Search completed: February 3, 2006, 21:56:33
Job time : 212.111 secs

15-JAN-2003; 2003US-00345837.
15-JAN-2003; 2003US-00345837.
(LINB/) LIN B.
Lin B;
WPI; 2004-517182/49.
P-PSDB; ADQ74869.
New substantially pure androgen responsive specific nucleic acid, useful
for diagnosing and treating prostate cancer.
Claim 148; SEQ ID NO 21; 102pp; English.
The invention relates to human androgen responsive prostate specific
(ARP) polynucleotides and the polypeptides they encode. The invention
also relates to a method of diagnosing or predicting susceptibility to a
prostate neoplastic condition in an individual and a method for treating
or reducing the severity of a prostate neoplastic condition in an
individual. The polynucleotides, polypeptides and methods of the
invention are useful for diagnosing and treating prostate cancer. This
sequence represents a human ARP polynucleotide of the invention.
Sequence 3318 BP; 842 A; 987 C; 828 G; 661 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 12; Length 3318;
Best Local Similarity 83.3%; Pred. No. 6.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACCAATGACCAACTGGGCTGT 25
Db 2802 AAATTCAATCACCACCACTGTGCTGT 2825

Search completed: February 3, 2006, 21:56:33
Job time : 212.111 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-5  
Perfect score: 25  
Sequence: 1 aaaaaacaatgacaaactgggctgt 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues  
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

- EST:\*
- 1: gb\_est1:\*
  - 2: gb\_est2:\*
  - 3: gb\_est3:\*
  - 4: gb\_est4:\*
  - 5: gb\_est5:\*
  - 6: gb\_est6:\*
  - 7: gb\_est7:\*
  - 8: gb\_est8:\*
  - 9: gb\_est9:\*
  - 10: gb\_est10:\*
  - 11: gb\_est11:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	327	2	BB204269
2	23.4	93.6	380	1	AI661876
3	23.4	93.6	526	1	AA403507
4	20.8	83.2	340	8	T85622
5	20.8	83.2	468	1	AA826673
6	20.8	83.2	539	6	CA439902
7	20.8	83.2	591	3	BM826284
8	20.8	83.2	727	6	CB241074
9	20.8	83.2	737	6	CD369726
10	20.8	83.2	747	5	BU616130
11	20.8	83.2	789	9	BH722326
12	20.8	83.2	803	2	EG741056
13	20.8	83.2	850	7	CR994650
14	20.8	83.2	1704	4	CR591670
15	20.2	80.8	556	1	AA596588
16	20.2	80.8	562	9	AQ279671
17	20.2	80.8	719	7	CO882131
18	19.8	79.2	373	8	H75316
19	19.8	79.2	451	7	CK650993
20	19.8	79.2	548	7	CK650237
21	19.8	79.2	683	6	CF178286
22	19.4	77.6	422	7	CK011470







Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 340)

#### REFERENCE AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, L., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE The WashU-Merck EST Project

#### JOURNAL

Unpublished (1995)

#### COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 899

High quality sequence stops: 242 Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 899 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 242.

#### FEATURES

source

Location/Qualifiers

1..340

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:467848"

/db\_xref="taxon:9606"

/clone="IMAGE:11231"

/sex="male"

/dev\_stage="20 week-post conception fetus"

/lab\_host="DH108 (ampicillin resistant)"

/clone\_lib="Soares fetal liver spleen INFILS"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)

with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer

15' AACTGGAGAAATTAATAAGATCTTTTCTTTTCTTTT 3',

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

#### ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 340;  
Best Local Similarity 91.7%; Pred. No. 1.9e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24

||||| ||||| ||||| ||||| |||||

Db 153 AAAAAACAATGTCACACTGGTCTG 176

#### RESULT 5

AA826673/c

#### LOCUS

of34g02.s1 NCI CGAP Kid6 Homo sapiens cDNA clone IMAGE:1423058 3'

similar to contains Alu repetitive element,, mRNA sequence.

#### DEFINITION

AA826673

#### ACCESSION

AA826673

#### VERSION

AA826673.1 GI:2900670

#### KEYWORDS

EST.

#### SOURCE

Homo sapiens

#### ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 468)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 861 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 387.

#### FEATURES

source

Location/Qualifiers

1..468

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1423058"

/sex="mixed"

/tissue\_type="kidney tumor"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="NCI CGAP Kid6"

/note="Organ: kidney; Vector: Bluescript SK-; Site 1:

EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'

GAATTCGCACGAG 3' 3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

#### ORIGIN

Query Match 83.2%; Score 20.8; DB 1; Length 468;  
Best Local Similarity 91.7%; Pred. No. 2e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24

||||| ||||| ||||| ||||| |||||

Db 280 AAAAAACAATGTCACACTGGTCTG 257

#### RESULT 6

CA439902/c

#### LOCUS

UI-H-D10-aux-k-21-0-UI.s1 NCI CGAP D10 Homo sapiens cDNA clone

DEFINITION UI-H-D10-aux-k-21-0-UI 3', mRNA sequence.

ACCESSION CA439902

VERSION CA439902.1 GI:24804322

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 539)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Jose Mercuende

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained

from Dr. M. Bento Soares, bento-soares@uiowa.edu

The following repetitive elements were found in this cDNA

sequence: 1-33, >AT rich#Low complexity (matched complement)

36-164, >ALU (matched complement)

Seq primer: M13 FORWARD

POLYA=yes.

Location/Qualifiers

1..539

/organism="Homo sapiens"

/mol\_type="mRNA"

source

/db\_xref="taxon:9606"  
 /clone="UI-H-D10-aux-21-0-UI"  
 /tissue\_type="Lung Focal Fibrosis"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP D10"  
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP D10 is a cDNA library containing the following tissue(s): A pool of Lung Focal Fibrosis. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (gt)18 tail. The sequence tag for this library is ATACGGGTC.  
 TAG\_TISSUE=lung with fibrosis  
 TAG\_LIB=UI-H-D10  
 TAG\_SEQ=ATACGGGTC"

## ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 539;  
 Best Local Similarity 91.7%; Pred. No. 2e+02; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCACTGGGCTG 24  
 |||||  
 Db 337 AAAAAACAATGTCACACTGGTCTG 314

## RESULT 7

BM826284 591 bp mRNA linear EST 06-MAR-2002  
 LOCUS K-EST0098334 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-106-D09  
 5', mRNA sequence.

ACCESSION BM826284  
 VERSION BM826284.1 GI:19182697  
 KEYWORDS EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 (bases 1 to 591)  
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.

## TITLE

21C Frontier Korean EST Project 2001

## JOURNAL

Unpublished (2002)

## COMMENT

Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongang@mail.kribb.re.kr  
 Plate: 106 row: D column: 09  
 High quality sequence stop: 591.

## FEATURES

## source

1. .591  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="S22SNU16n1-106-D09"  
 /sex="F"  
 /tissue\_type="Ascites"  
 /cell\_type="Lymphoblast-like"  
 /cell\_line="SNU-16"  
 /lab\_host="DH10B"

/clone\_lib="S22SNU16n1"  
 /note="Organ: Stomach; Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI; The S22SNU16 library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from harvested cells of SNU-16 culture. SNU-16 cell was obtained from Korean Cell Line Bank (KCLB). SNU-16 was established from ascitic fluids of Korean patients by Park J.G. et al. (1990). Cancer Res 50: 2773-2780."

## ORIGIN

Query Match 83.2%; Score 20.8; DB 3; Length 591;  
 Best Local Similarity 91.7%; Pred. No. 2e+02; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCACTGGGCTG 24  
 |||||  
 Db 60 AAAAAACAATGTCACACTGGTCTG 83

## RESULT 8

CB241074/c

LOCUS

DEFINITION

CB241074

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 727)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).

The following repetitive elements were found in this cDNA

sequence: 1-23, >AT richLow complexity 315-341,

>AT richLow complexity 344-472, >ALU (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

## FEATURES

## source

1. .727  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-FNO-afx-m-14-0-UI"  
 /tissue\_type="Human Lung Epithelial cells"  
 /lab\_host="DH10B (Life Technologies)" (T1 phage resistant)"  
 /clone\_lib="UI-CF-FNO"  
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-CF-FNO is a subtracted cDNA library derived from two normalized Human lung epithelial cell libraries (EN1 and DUI). The library was subtracted according to according to



modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP.DF0 is a cDNA library containing the following tissue(s): Subchondral Bone. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTVT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTACGCTC.  
TAG\_TISSUE=subchondral bone  
TAG\_LiB=UI-H-DF0  
TAG\_SEQ=GTTACGCTC"

## ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 747;  
Best Local Similarity 91.7%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTG 24  
|||||  
Db 651 AAAAAAATGTCCTCACTGCTG 628  
|||||

## RESULT 11

BH722326/c  
LOCUS BH722326 789 bp DNA linear GSS 20-FEB-2002  
DEFINITION BOMGS07TR BO\_2\_3 KB Brassica oleracea genomic clone BOMGS07,  
genomic survey sequence.

ACCESSION BH722326  
VERSION BH722326.1 GI:18825399  
SOURCE GSS.  
ORGANISM Brassica oleracea

Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 789)  
Ayele, M., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,  
Utterback, T.R., Wortman, J.R., White, O.R. and Town, C.D.

Whble genome shotgun sequencing of Brassica oleracea and its  
application to gene discovery and annotation in Arabidopsis  
Genome Res. 15 (4), 487-495 (2005)  
15805490  
Other GSSs: BOMGS07TF  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR  
Class: sheared ends.

Location/Qualifiers

1..789

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="TO100DH3"

/db\_xref="taxon:3712"

/clone\_lib="BOMGS07"

/clone\_lib="BO\_2\_3\_KB"

/note="Vector: pBOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pBOS1 using BstXI linkers"

## FEATURES

source

Query Match 83.2%; Score 20.8; DB 9; Length 789;  
Best Local Similarity 91.7%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## ORIGIN

QY 1 AAAAAAATGACCAACTGGGCTG 24  
|||||

Db 392 AAAAAAATGACCAACTGGGTTG 369  
|||||

## RESULT 12

BG741056  
LOCUS BG741056 803 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602634789F1 NCI\_CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4779574 5',  
mRNA sequence.

ACCESSION BG741056

VERSION BG741056.1 GI:14051709

SOURCE EST.

KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 803)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: James Cleaver, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10637 row: b column: 23  
High quality sequence stop: 697.

Location/Qualifiers

1..803

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4779574"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI CGAP\_Skn3"

/note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 803;  
Best Local Similarity 91.7%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTG 24  
|||||

Db 406 AAAAAAATGTCCTCACTGCTG 429  
|||||

## RESULT 13

CR994650  
LOCUS CR994650 850 bp mRNA linear EST 28-JUN-2005  
DEFINITION CR994650 RZPD no.9016 Homo sapiens cDNA clone RZPDp9016M1729 5',  
mRNA sequence.

ACCESSION CR994650

VERSION CR994650.1 GI:68288535

SOURCE EST.

KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 850)  
Heil, O., Ebert, L., Hennig, S., Henze, S., Radelof, U., Schneider, D.  
Human T-Lymphocytes library  
Unpublished (2005)

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 789;  
Best Local Similarity 91.7%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTG 24  
|||||

Db 406 AAAAAAATGTCCTCACTGCTG 429  
|||||

```

COMMENT
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; RZPDp9016M1729.
RZPDLIB; (Human T-Lymphocytes) RZPD LIB No.9016
http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=9016 Contact:
Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available from RZPD:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPDp9016M1729
contact RZPD (product- support@rzpd.de) for further information.
Primer name: q3.4 , Primer sequence: CGGATAACAATTTCACACAG.
FEATURES
Location/Qualifiers
1..850
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="RZPDp9016M1729"
/tissue_type="T-Lymphocytes"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RZPD no.9016"
/note="Vector: pQE80LSN_cloned; Site_1: SalI; Site_2:
NotI; vector:
http://www.rzpd.de/info/vectors/pQE80LSN_cloned.pic.shtml
; 1st strand cDNA was prepared from mRNA obtained from
human T-Lymphocytes with a NotI - oligo(dT) primer [5',
GACTAGTCTAGATCGGAGCGCCGCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to SalI adaptors,
digested with NotI and cloned into the NotI and SalI sites
of the pQE80LSN_cloned vector"
ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 850;
Best Local Similarity 91.7%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
|||||
Db 363 AAAAAACAATGACCAACTGGTCTG 386

RESULT 14
CR591670/c
LOCUS
DEFINITION
full-length cDNA clone CS0DF030YN16 of Fetal brain of Homo sapiens
(human).
ACCESSION
CR591670.1 GI:50472477
VERSION
HTC; CnSLT cDNA.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 1704)
AUTHORS
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1704)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
Bp 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; RZPDp9016M1729.
RZPDLIB; (Human T-Lymphocytes) RZPD LIB No.9016
http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=9016 Contact:
Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available from RZPD:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPDp9016M1729
contact RZPD (product- support@rzpd.de) for further information.
Primer name: q3.4 , Primer sequence: CGGATAACAATTTCACACAG.
FEATURES
Location/Qualifiers
1..1704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF030YN16"
/tissue_type="Fetal brain"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 83.2%; Score 20.8; DB 4; Length 1704;
Best Local Similarity 91.7%; Pred. NO. 2.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
|||||
Db 283 AAAAAACAATGACCAACTGGTCTG 260

RESULT 15
AA596588
LOCUS
DEFINITION
vms8f12.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
IMAGE:1002479 5', mRNA sequence.
ACCESSION
AA596588
VERSION
AA596588.1 GI:2412023
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 556)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra W/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:566695
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 442.
Location/Qualifiers
1..556
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1002479"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse Tcell 937311"
/note="Organ: blood; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GATTCGCGCAGG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
ORIGIN

```

```

COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF030YN16"
/tissue_type="Fetal brain"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 83.2%; Score 20.8; DB 4; Length 1704;
Best Local Similarity 91.7%; Pred. NO. 2.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
|||||
Db 283 AAAAAACAATGACCAACTGGTCTG 260

RESULT 15
AA596588
LOCUS
DEFINITION
vms8f12.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
IMAGE:1002479 5', mRNA sequence.
ACCESSION
AA596588
VERSION
AA596588.1 GI:2412023
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 556)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra W/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:566695
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 442.
Location/Qualifiers
1..556
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1002479"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse Tcell 937311"
/note="Organ: blood; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GATTCGCGCAGG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
ORIGIN

```

```

Query Match      80.8%; Score 20.2; DB 1; Length 556;
Best Local Similarity 88.0%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 373 AAGAAACAAGACCAACTGAGCTGT 397

RESULT 16
LOCUS AQ279671/c
DEFINITION CITBI-E1-2513L17.TR CITBI-E1 Homo sapiens genomic clone 2513L17,
genomic survey sequence.
ACCESSION AQ279671
VERSION AQ279671.1 GI:3905575
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchotheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 562)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..562
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2513L17"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /notes="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"

ORIGIN
Query Match      80.8%; Score 20.2; DB 9; Length 562;
Best Local Similarity 88.0%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 435 AAAAAACAATGATATCTGGGCTGT 411

RESULT 17
LOCUS CO892131/c
DEFINITION Bovgen_20456 normal cattle brain Bos taurus cDNA clone
RZPDp1056D2337Q 5', mRNA sequence.
ACCESSION CO892131
VERSION CO892131.1 GI:51822432
KEYWORDS EST.
SOURCE Bos taurus (cow)

Query Match      80.8%; Score 20.2; DB 7; Length 719;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 162 AAAAAACAATGATATCTGGGCTGT 138

RESULT 18
LOCUS H75316
DEFINITION Yu06h02.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:233043 5', mRNA sequence.
ACCESSION H75316
VERSION H75316.1 GI:1049739
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 373)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favallo,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,B., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 719)
AUTHORS Hennig,S., Janitz,M., Herwig,R. and Williams,J.
TITLE Generation, annotation, evolutionary analysis and database
integration of 14969 cattle EST clusters
JOURNAL Unpublished (2004)
COMMENT Contact: Hennig S
laboraty 123, dept.Lehbrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones that display the same hybridisation matrix with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per ONFP cluster was selected for sequencing. cDNA clones and
filters are distributed via Deutsches Ressourcenzentrum fuer
Genomforschung GmbH (http://www.rzpd.de).
PCR Primers
FORWARD: 5' CCCCGAGCTTTACACTTTATGCTCCGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTACGCAGCTGGCGAAGGGGATGTG 3' (M13FSP) 3'-seq
Seq primer: 5'-CCGTCGCGAATTCCCGGCT-3' (M13RSP).

FEATURES             Location/Qualifiers
     source           1..719
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="RZPDp1056D2337Q"
                     /sex="female"
                     /tissue_type="brain tissue"
                     /dev_stage="adult brain"
                     /clone_lib="normal cattle brain"
                     /note="Organ: brain; Vector: pSport1; Site 1: NotI;
                     Site 2: Sali; Random primed and directionally cloned in
                     pSport1 vector using NotI
                     (5'-pGACTAGTCTAGATCGGACGGCGGCC (T)15-3' and Sali 5'-
                     TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"

ORIGIN
Query Match      80.8%; Score 20.2; DB 7; Length 719;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 162 AAAAAACAATGATATCTGGGCTGT 138

```

TITLE  
JOURNAL  
PUBMED  
COMMENT

and Maria,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
889549  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 857  
High quality sequence stops: 313  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 857 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 313.

FEATURES  
source  
1..373  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3786220"  
/db\_xref="taxon:9606"  
/clone="IMAGE:233043"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
15' AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3',  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 79.2%; Score 19.8; DB 8; Length 373;  
Best Local Similarity 91.3%; Pred. No. 5.4e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 146 AAAACAATGACCAACTGGCTG 168  
|||||  
RESULT 19  
CK650993 451 bp mRNA linear EST 31-DEC-2004  
LOCUS  
DEFINITION  
m.03.E12.5 MPer183 cassava lambda zap Manihot esculenta cDNA 5',  
mRNA sequence.  
ACCESSION  
CK650993.1 GI:56926656  
VERSION  
EST.  
KEYWORDS  
SOURCE  
Manihot esculenta (cassava)  
ORGANISM  
Manihot esculenta  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Malpighiales; Euphorbiales; Crotonoideae;  
Manihoteae; Manihot.  
1 (bases 1 to 451)  
Lopez,C., Jorge,V., Piegue,B., Mba,C., Cortes,D., Restrepo,S.,  
Soto,M., Laudie,M., Berger,C., Cooke,R., Delsey,M., Tohme,J. and  
Verdier,V.  
A unigene catalogue of 5700 expressed genes in cassava  
Plant Mol. Biol. 56 (4), 541-554 (2004)  
Contact: Verdier V  
Laboratoire genome et developpement des plantes, UMR5096  
IRD-CNRS-Universite de Perpignan-CIAT

FEATURES  
source  
1..548  
/organism="Manihot esculenta"  
/mol\_type="mRNA"  
/cultivar="MPer183"  
/db\_xref="taxon:3983"  
/clone\_lib="MPer183 cassava lambda zap"  
/note="Organ: Root; Vector: pBluescript SK (+/-); Site\_1:  
EcoRI; Site\_2: XhoI; Stratagene lambda Zap Library"

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

52 AV Paul Alduy Perpignan 66860 France  
Tel: (33) 4 68 66 17 74  
Fax: (33) 4 68 66 84 99  
Email: Valerie.Verdier@ird.fr  
Seq primer: T3.  
Location/Qualifiers  
1..451  
/organism="Manihot esculenta"  
/mol\_type="mRNA"  
/cultivar="MPer183"  
/db\_xref="taxon:3983"  
/clone\_lib="MPer183 cassava lambda zap"  
/note="Organ: Root; Vector: pBluescript SK (+/-); Site\_1:  
EcoRI; Site\_2: XhoI; Stratagene lambda Zap Library"

FEATURES  
source  
1..451  
/organism="Manihot esculenta"  
/mol\_type="mRNA"  
/cultivar="MPer183"  
/db\_xref="taxon:3983"  
/clone\_lib="MPer183 cassava lambda zap"  
/note="Organ: Root; Vector: pBluescript SK (+/-); Site\_1:  
EcoRI; Site\_2: XhoI; Stratagene lambda Zap Library"

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 451;  
Best Local Similarity 91.3%; Pred. No. 5.5e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 300 AAAACAATGACCAACTGGCTG 322  
|||||

RESULT 20  
CK650237 548 bp mRNA linear EST 31-DEC-2004  
LOCUS  
DEFINITION  
m.11.L23.5 MPer183 cassava lambda zap Manihot esculenta cDNA 5',  
mRNA sequence.  
ACCESSION  
CK650237  
VERSION  
EST.  
KEYWORDS  
SOURCE  
Manihot esculenta (cassava)  
ORGANISM  
Manihot esculenta  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Malpighiales; Euphorbiales; Crotonoideae;  
Manihoteae; Manihot.  
1 (bases 1 to 548)  
Lopez,C., Jorge,V., Piegue,B., Mba,C., Cortes,D., Restrepo,S.,  
Soto,M., Laudie,M., Berger,C., Cooke,R., Delsey,M., Tohme,J. and  
Verdier,V.  
A unigene catalogue of 5700 expressed genes in cassava  
Plant Mol. Biol. 56 (4), 541-554 (2004)  
Contact: Verdier V  
Laboratoire genome et developpement des plantes, UMR5096  
IRD-CNRS-Universite de Perpignan-CIAT

FEATURES  
source  
1..548  
/organism="Manihot esculenta"  
/mol\_type="mRNA"  
/cultivar="MPer183"  
/db\_xref="taxon:3983"  
/clone\_lib="MPer183 cassava lambda zap"  
/note="Organ: Root; Vector: pBluescript SK (+/-); Site\_1:  
EcoRI; Site\_2: XhoI; Stratagene lambda Zap Library"

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||

ORIGIN  
Query Match 79.2%; Score 19.8; DB 7; Length 548;  
Best Local Similarity 91.3%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAACAATGACCAACTGGCTG 24  
|||||  
Db 435 AAAACAATGACCAACTGGCTG 457  
|||||



```

RESULT 21
CF178286/c
LOCUS       CF178286               683 bp    mRNA    linear    EST 28-JUL-2003
DEFINITION   807338 MARC 3PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   CF178286
VERSION     CF178286.1   GI:33290062
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE   1 (bases 1 to 683)
AUTHORS     Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
            Wise,I.A., Nonneman,B.J., Wray,J.E. and Keele,J.W.
TITLE       A second set of porcine ESTs from a pooled-tissue normalized
            library
JOURNAL     Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            FO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: SRG8006 row: B column: 22
            Seq primer: GTAATACGACTCTACTATAGG.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3PIG"
                     /notes="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including brain, liver, muscle, placenta/endometrium,
                     ovary, testes, and bone marrow."
ORIGIN
Query Match       79.2%;   Score 19.8;   DB 6;   Length 683;
Best Local Similarity 91.3%;   Pred. No. 5.8e+02;
Matches 21;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

QY       1  AAAAAACATGACCAACTGGGCT 23
          |||||  |||||  |||||  |||||  |||||
Db       323  AAAAAACAGGACCAACTGGGCT 301

RESULT 22
CK011470/c
LOCUS       CK011470               422 bp    mRNA    linear    EST 04-FEB-2005
DEFINITION   29146rsicenf_7286.y1 Oryza sativa cv. PA64s panicle sterile cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK011470
VERSION     CK011470.1   GI:58600942
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM    Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 422)
AUTHORS     Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
TITLE       The Genomes of Oryza sativa: A History of Duplications
JOURNAL     PLoS Biol. 3 (2), e38 (2005)
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 662
            POLYA=No.
            Location/Qualifiers
     source          1..422
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="panicle"
                     /cell_type="sterile"
                     /dev_stage="heading/flowering"
                     /clone_lib="Oryza sativa cv. PA64s panicle sterile cDNA
                     library"
ORIGIN
Query Match       77.6%;   Score 19.4;   DB 7;   Length 422;
Best Local Similarity 95.2%;   Pred. No. 8.3e+02;
Matches 20;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY       3  AAAACAATGACCAACTGGGCT 23
          |||||  |||||  |||||  |||||  |||||
Db       123  AAAATAATGACCAACTGGGCT 103

RESULT 23
CK063363/c
LOCUS       CK063363               662 bp    mRNA    linear    EST 05-FEB-2005
DEFINITION   73210rsicem_6604.y1 Oryza sativa cv. PA64s panicle sterile cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK063363
VERSION     CK063363.1   GI:58674676
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM    Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 662)
AUTHORS     Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
TITLE       The Genomes of Oryza sativa: A History of Duplications
JOURNAL     PLoS Biol. 3 (2), e38 (2005)
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 662
            POLYA=No.
            Location/Qualifiers
     source          1..422
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="panicle"
                     /cell_type="sterile"
                     /dev_stage="heading/flowering"
                     /clone_lib="Oryza sativa cv. PA64s panicle sterile cDNA
                     library"
ORIGIN
Query Match       77.6%;   Score 19.4;   DB 7;   Length 422;
Best Local Similarity 95.2%;   Pred. No. 8.3e+02;
Matches 20;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY       3  AAAACAATGACCAACTGGGCT 23
          |||||  |||||  |||||  |||||  |||||
Db       123  AAAATAATGACCAACTGGGCT 103

RESULT 23
CK063363/c
LOCUS       CK063363               662 bp    mRNA    linear    EST 05-FEB-2005
DEFINITION   73210rsicem_6604.y1 Oryza sativa cv. PA64s panicle sterile cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK063363
VERSION     CK063363.1   GI:58674676
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM    Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 662)
AUTHORS     Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
TITLE       The Genomes of Oryza sativa: A History of Duplications
JOURNAL     PLoS Biol. 3 (2), e38 (2005)
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 662
            POLYA=No.
            Location/Qualifiers
     source          1..422
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="panicle"
                     /cell_type="sterile"
                     /dev_stage="heading/flowering"
                     /clone_lib="Oryza sativa cv. PA64s panicle sterile cDNA
                     library"
ORIGIN
Query Match       77.6%;   Score 19.4;   DB 7;   Length 422;
Best Local Similarity 95.2%;   Pred. No. 8.3e+02;
Matches 20;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY       3  AAAACAATGACCAACTGGGCT 23
          |||||  |||||  |||||  |||||  |||||
Db       123  AAAATAATGACCAACTGGGCT 103

```



**TITLE** Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction  
**JOURNAL** Plant Physiol. 138 (1), 105-115 (2005)  
**PUBMED** 15888683  
**COMMENT** Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: rwing@genome.arizona.edu

PCR Primers  
 FORWARD: gta aaa cga cgg cca gtg  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 15 row: D column: 19  
 Seq primer: gga aac agc tat gac cat g.

**FEATURES**  
**source**  
 1. .812  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNEe15D19"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEe"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**

Query Match 77.6%; Score 19.4; DB 6; Length 812;  
 Best Local Similarity 95.2%; Pred. No. 8.9e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 3 AAAACAATGACCAACTGGGCT 23  
 |||||  
**Db** 775 AAAATAATGACCAACTGGGCT 795

**RESULT 27**  
**CB677843**  
**LOCUS** OSJNEe15D18 r OSJNEe Oryza sativa (japonica cultivar-group) cDNA  
**DEFINITION** clone OSJNEe15D18 3', mRNA sequence.  
**ACCESSION** CB677843  
**VERSION** CB677843.1 GI:29681568  
**KEYWORDS** EST.  
**SOURCE** Oryza sativa (japonica cultivar-group)  
**ORGANISM** Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

**REFERENCE**  
**AUTHORS** Jantaseuriyarat, C., Gowda, M., Haller, K., Hatfield, J., Lu, G., Stahlberg, E., Zhou, B., Li, H., Kim, H., Yu, Y., Dean, R.A., Wing, R.A., Soderlund, C. and Wang, G.L.  
**TITLE** Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction  
**JOURNAL** Plant Physiol. 138 (1), 105-115 (2005)  
**PUBMED** 15888683  
**COMMENT** Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: rwing@genome.arizona.edu

PCR Primers  
 FORWARD: gta aaa cga cgg cca gtg  
 BACKWARD: gga aac agc tat gac cat g

**FEATURES**  
**source**  
 1. .911  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNEe15D18"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEe"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**

Query Match 77.6%; Score 19.4; DB 6; Length 911;  
 Best Local Similarity 95.2%; Pred. No. 9.1e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 3 AAAACAATGACCAACTGGGCT 23  
 |||||  
**Db** 828 AAAATAATGACCAACTGGGCT 848

**RESULT 28**  
**CF330489**  
**LOCUS** NACL--06-D15.g1 Rice callus plasmid cDNA library (NACL) Oryza sativa (japonica cultivar-group) cDNA clone NACL--06-D15, mRNA sequence.  
**DEFINITION** CF330489.1 GI:33809212  
**ACCESSION** CF330489  
**VERSION** CF330489.1  
**KEYWORDS** EST.  
**SOURCE** Oryza sativa (japonica cultivar-group)  
**ORGANISM** Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

**REFERENCE**  
**AUTHORS** Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C., Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.  
**TITLE** Large-scale Sequencing Analysis of Rice ESTs  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Nahm B.H.  
 Genomics and Genetics Institute, GreenGene Biotech Inc.; Division of Bioscience and Bioinformatics, Myongji University  
 Yongin, Kyeonggi, Korea  
 Tel: 82 31 330 6193  
 Fax: 82 31 321 6355  
 Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

**FEATURES**  
**source**  
 1. .163  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nackdong"  
 /db\_xref="taxon:39947"  
 /clone="NACL--06-D15"  
 /tissue\_type="callus"  
 /dev\_stage="proliferated callus on 2N6 media for 30 days"  
 /lab\_host="E.coli DH10B"  
 /clone\_lib="Rice callus plasmid cDNA library (NACL)"  
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; mRNA was capped with oligoribonucleotides and then used as templates for RT-PCR."

**ORIGIN**

Query Match 76.8%; Score 19.2; DB 6; Length 163;  
 Best Local Similarity 87.5%; Pred. No. 9.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 2 AAAACAATGACCAACTGGGCTGT 25

```

Db      69 AAAAAAGTGACCAACTGGGCTGT 92
|||||  |||||||  |||||||  |||||||  |||||||
RESULT 29
CF330488/c
LOCUS   CF330488.1 192 bp mRNA linear EST 18-AUG-2003
DEFINITION NACL--06-D15.b1 Rice callus plasmid cDNA library (NACL) Oryza
           sativa (japonica cultivar-group) cDNA clone NACL--06-D15, mRNA
           sequence.
ACCESSION CF330488
VERSION   CF330488.1 GI:33809208
KEYWORDS EST.
SOURCE   Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 192)
AUTHORS   Kim,J.S., Jun,K.W., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
          Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.
TITLE     Large-scale Sequencing Analysis of Rice ESTs
JOURNAL   Unpublished (2003)
COMMENT   Contact: Nahm B.H.
          Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
          of Bioscience and Bioinformatics, Myongji University
          Yongin, Kyeonggi, Korea
          Tel: 82 31 330 6193
          Fax: 82 31 321 6355
          Email: bhnaemggbio.com, bhnaem@bio.myongji.ac.kr.

FEATURES             source
1..192
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="NACL--06-D15"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 30 days"
/lab_host="E.coli DH10B"
/clone_lib="Rice callus plasmid cDNA library (NACL)"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; mRNA was capped
with oligoribonucleotides and then used as templates for
RT-PCR."

ORIGIN
Query Match      76.8%; Score 19.2; DB 6; Length 192;
Best Local Similarity 87.5%; Pred. No. 9.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAAACAATGACCAACTGGGCTGT 25
|||||  |||||||  |||||||  |||||||  |||||||
Db      95 AAAAAAGTGACCAACTGGGCTGT 72
|||||  |||||||  |||||||  |||||||  |||||||

RESULT 30
CL605612/c
LOCUS   CL605612.1 471 bp DNA linear GSS 17-JUN-2004
DEFINITION CH240_182B22.TV CHORI-240 Bos taurus genomic clone CH240_182B22,
           genomic survey sequence.
ACCESSION CL605612
VERSION   CL605612.1 GI:48973644
KEYWORDS GSS.
SOURCE   Bos taurus (cow)
ORGANISM Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
          Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 471)
AUTHORS   Costa,J.N., Mota,M. and Caetano,A.R.
TITLE     Brazil's Contribution to End-Sequencing the Bovine BAC Library
JOURNAL   CHORI-240
          Unpublished (2003)

Other GSSs: CH240_182B22.TV
Contact: Caetano AR
Department of Biotechnology
Embrapa Recursos Geneticos e Biotecnologia
Parque Estacao Biologica, Final Av. W/5 Norte, Brasilia-DF C.P.
02372, 70770-900 Brasil
Tel: 55 61 448 4778
Fax: 55 61 340 1658
Email: acetanoc@embrapa.br
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm).
Bases shown have phred quality value equal to or higher than 20.
Bases with quality value below 20 were masked with 'N'.
For BAC library availability, please contact Pieter de Jong
(pdejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm).
This work was undertaken as part of the International Bovine BAC
Mapping Consortium (IBBMC) by Embrapa Recursos Geneticos e
Biotecnologia with financing from Conselho Nacional de
Desenvolvimento Cientifico e Tecnologico (CNPq), Brazil
Plate: 182 row: B column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 471.
Location/Qualifiers
1..471
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_182B22"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match      76.8%; Score 19.2; DB 10; Length 471;
Best Local Similarity 87.5%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAACAATGACCAACTGGGCTG 24
|||||  |||||||  |||||||  |||||||  |||||||
Db      334 AAAAAACAATGATCAAGTGGACTG 311
|||||  |||||||  |||||||  |||||||  |||||||

RESULT 31
AZ307205/c
LOCUS   AZ307205.1 472 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0008D11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
           clone UUGC1M0008D11 R, genomic survey sequence.
ACCESSION AZ307205
VERSION   AZ307205.1 GI:10345973
KEYWORDS GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 472)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Becorn,T., Duval,B., Hamil,C.,
          Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
          Reilly,M., Rose,R., Rose,R., Stokes,R., Tingey,A., von
          Niederhausern,A. and Wright,D., Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
           plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
          University of Utah Genome Center
          University of Utah

```



/clone\_lib="Rice panicle at flowering stage"  
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 709;  
Best Local Similarity 87.5%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTGT 25

Db 631 AAAAAAGTGTGACCAACTGGGCTGT 654

## RESULT 34

CX248961/c

LOCUS 1301028 NCCWA 02RT Oncorhynchus mykiss cDNA 3', mRNA EST 01-MAR-2005  
DEFINITION CX248961

ACCESSION

VERSION CX248961.1 GI:60365483

KEYWORDS

SOURCE EST.

ORGANISM Oncorhynchus mykiss (rainbow trout)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE 1 (bases 1 to 719)

AUTHORS Rexroad, C.E., Goupil, A.-S., Guiguen, Y. and Yao, J.

TITLE 02RT IUS, NCCWA/WVU EST Project, Phase II, in collaboration with

JOURNAL INRA

COMMENT Unpublished (2004)

Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim\_alt option. Vector identified with

cross\_match v0.990329.

Plate: 106 row: F column: 4

Seq primer: GTAATACGACTCACTATAGG.

## FEATURES

source

Location/Qualifiers

1..719

/organism="Oncorhynchus mykiss"

/mol\_type="mRNA"

/db\_xref="taxon:8022"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="NCCWA 02RT"

/note="Vector: pCMV Sport6.0; This library was created by  
A.-S. Goupil and Y. Guiguen who subcloned the NCCWA 1RT  
library from the INRA multi-tissue library."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 719;  
Best Local Similarity 87.5%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTG 24

Db 287 AAAAAAAGATGAACCTGGGCTG 264

## RESULT 35

CR058845/c

LOCUS 779 bp DNA linear GSS 05-JUL-2004

DEFINITION Reverse strand read from insert in 3'HPRT insertion targeting and

chromosome engineering clone MHPP304h18, genomic survey sequence.

ACCESSION CR058845

VERSION CR058845.1 GI:49792317

KEYWORDS GSS; genome survey sequence; MICER.

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Murinae; Mus.

REFERENCE 1 (bases 1 to 779)

ADAMS, D.J., BIGGS, P.J., COX, A.V., DAVIES, R.M., VAN DER WEYDEN, L.,

JONKERS, J., SMITH, J., PLUMB, R.W., TAYLOR, R.G., NISHIJIMA, I., YU, Y.,

ROGERS, J. and BRADLEY, A.

Direct Submission

Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. http://www.sanger.ac.uk/MICER

## FEATURES

source

Location/Qualifiers

1..779

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="MHPP304h18"

/clone\_lib="MHPP"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 779;  
Best Local Similarity 87.5%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTGT 25

Db 72 AAAAAAATGACCAACTGGGCTGT 49

## RESULT 36

CR021797/c

LOCUS 792 bp DNA linear GSS 05-JUL-2004

DEFINITION Reverse strand read from insert in 3'HPRT insertion targeting and

chromosome engineering clone MHPP270a06, genomic survey sequence.

ACCESSION CR021797

VERSION CR021797.1 GI:49754852

KEYWORDS GSS; genome survey sequence; MICER.

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Murinae; Mus.

REFERENCE 1 (bases 1 to 792)

ADAMS, D.J., BIGGS, P.J., COX, A.V., DAVIES, R.M., VAN DER WEYDEN, L.,

JONKERS, J., SMITH, J., PLUMB, R.W., TAYLOR, R.G., NISHIJIMA, I., YU, Y.,

ROGERS, J. and BRADLEY, A.

Direct Submission

Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. http://www.sanger.ac.uk/MICER

## FEATURES

source

1..792

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="MHPP270a06"

/clone\_lib="MHPP"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 792;  
Best Local Similarity 87.5%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTGT 25

Db 72 AAAAAAATGACCAACTGGGCTGT 49

## RESULT 37

CR090689/c

LOCUS 801 bp DNA linear GSS 05-JUL-2004

DEFINITION Reverse strand read from insert in 3'HPRT insertion targeting and

chromosome engineering clone MHPP304g17, genomic survey sequence.

ACCESSION CR090689

SOURCE Mus musculus (house mouse)

```
VERSION CR090689.1 GI:49824440
KEYWORDS GSS: genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES
source
1..801
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPP304G17"
/clone_lib="MHPP"
ORIGIN
Query Match 76.8%; Score 19.2; DB 11; Length 801;
Best Local Similarity 87.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAACAATGACCAACTGGGCTGT 25
|||||
Db 72 AAAAAACAATGACCAAGCGGTTGT 49
|||||
RESULT 38
AG479423/c 808 bp DNA linear GSS 22-DEC-2004
LOCUS Mus musculus molossinus DNA, clone:MSMg01-373G01.TJ, genomic survey
DEFINITION sequence.
ACCESSION AG479423
VERSION AG479423.1 GI:48186653
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and
Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 808)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Taikuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY : pBACe3.6
Vector
```

```

R.Site 1 : EcoRI.
R.Site 2 : EcoRI.
Location/Qualifiers
1..808
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-373G01.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
Query Match 76.8%; Score 19.2; DB 10; Length 808;
Best Local Similarity 87.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACAATGACCAACTGGGCTG 24
|||||
Db 528 AAAAAAATATCACCACCTGGGCTG 505
|||||
RESULT 39
AU252333 811 bp mRNA linear EST 26-JAN-2005
LOCUS AU252333 salt-stressed barley root cDNA Hordeum vulgare subsp.
DEFINITION vulgare cDNA clone BR-T08 similar to HvLRR2, mRNA sequence.
ACCESSION AU252333
VERSION AU252333.1 GI:19494120
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 811)
AUTHORS Ueda,A., Shi,W., Nakamura,T. and Takabe,T.
TITLE Analysis of salt-inducible genes in barley roots by differential
display
JOURNAL J. Plant Res. 115 (1118), 119-130 (2002)
PUBMED 12884135
COMMENT Contact: Akihiro Ueda
Lab. of Biosphere Symbioses
Graduate School of Bioagricultural Sciences, Nagoya University
Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan
Tel: 81-52-789-4044(ex.5209)
Fax: 81-52-789-5209
E-mail: akihiro@nuagri.agr.nagoya-u.ac.jp.
FEATURES
source
1..811
Location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Haruna-ni-jo"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="BR-T08"
/tissue_type="root"
/dev_stage="early vegetative stage"
/clone_lib="salt-stressed barley root cDNA"
ORIGIN
Query Match 76.8%; Score 19.2; DB 1; Length 811;
Best Local Similarity 87.8%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACAATGACCAACTGGGCTG 24
|||||
Db 249 AAGATAAATGACCAACTGGGCTG 272
|||||
RESULT 40
CR101220/c
```





EST.  
SOURCE Homo sapiens (human)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 990)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 15, 2003 this sequence version replaced gi:30771985.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
This sequence belongs to sequence cluster 3595.r  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0DF023AC03QPL6c=3595.r.

#### FEATURES

Location/Qualifiers  
1..990  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF023YE05"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."  
.

#### ORIGIN

Query Match 76.0%; Score 19; DB 5; Length 990;  
Best Local Similarity 76.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCACTGGCGTGT 25  
|||||:::|||||||

Db 747 AAAAAAAAWMGCACTGGCGGT 771

RESULT 44  
LOCUS CNS00LH5 1101 bp DNA linear GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC: BACR30106 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL068211  
VERSION AL068211.1 GI:4958428  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw 9p, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

#### FEATURES

Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR30106"  
/clone\_lib="RPI-98"  
/note="end : 17"

#### ORIGIN

Query Match 76.0%; Score 19; DB 10; Length 1101;  
Best Local Similarity 76.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCACTGGCGTGT 25  
|||||:::|||||||

Db 450 AAAAAAAACCACTGGCGTGT 474

#### RESULT 45

LOCUS BW779425/c 333 bp mRNA linear EST 10-AUG-2005  
DEFINITION BW779425 Amphioxus Branchiostoma floridae unpublished cDNA library, gastrula whole animal Branchiostoma floridae cDNA clone bbg0055a01 5', mRNA sequence.

ACCESSION BW779425  
VERSION BW779425.1 GI:66381919  
KEYWORDS EST.  
SOURCE Branchiostoma floridae (Florida lancelet)  
ORGANISM Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

REFERENCE 1 (bases 1 to 333)  
AUTHORS Yu, J., Holland, L.Z., Shin-i, T., Kohara, Y., Satou, Y. and Satoh, N.  
TITLE Expressed genes in Branchiostoma floridae  
JOURNAL Unpublished (2005)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshinigenes.nig.ac.jp

If you want to have a cDNA clone for this EST or if you have any questions, please send an e-mail to Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

#### FEATURES

Location/Qualifiers  
1..333  
/organism="Branchiostoma floridae"  
/mol\_type="mRNA"  
/db\_xref="taxon:7739"  
/clone="bbg0055a01"  
/tissue\_type="whole animal"  
/dev\_stage="gastrula"  
/clone\_lib="Amphioxus Branchiostoma floridae unpublished cDNA library, gastrula whole animal"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 333;  
Best Local Similarity 87.0%; Pred. No. 1.5e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1920051C22"
/tissue_type="stomach"
/dev_stage="17 days embryo"
/clone_lib="RIKEN full-length enriched, 17 days embryo
stomach"

ORIGIN
Query Match 75.2%; Score 18.8; DB 5; Length 410;
Best Local Similarity 90.9%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACAAATGACCAACTGGGCTGT 25
||||| ||||| ||||| |||||
Db 357 AACAAAGGACCAATTTGGGCTGT 378

RESULT 48
AQT36236 511 bp DNA linear GSS 15-JUL-1999
LOCUS HS 3043_B2_P05_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3043 Col=10 Row=H, genomic survey
sequence.
ACCESSION AQT36236
VERSION AQT36236.1 GI:5507788
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 511)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 1049764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3043 row: H column: 10
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 511.
FEATURES
source
1..511
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3043 Col=10 Row=H"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"

ORIGIN
Query Match 75.2%; Score 18.8; DB 9; Length 511;
Best Local Similarity 90.9%; Pred. No. 1.6e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACAATGACCAACTGGGC 22
||||| ||||| ||||| |||||

Db 150 AAAAATAATGACCACTGGGC 171

RESULT 49
CB334117/c 523 bp mRNA linear EST 29-OCT-2003
LOCUS Mg_AFT_30D09_M13F Mesobuthus gibbosus adult female tail Mesobuthus
DEFINITION gibbosus cDNA clone Mg_AFT_30D09 5' similar to P91468 (P91468)
T20D4.7 protein, mRNA sequence.
ACCESSION CB334117
VERSION CB334117.1 GI:28915869
KEYWORDS EST.
SOURCE Mesobuthus gibbosus (Mediterranean checkered scorpion)
ORGANISM Mesobuthus gibbosus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
Buthida; Buthoidea; Buthidae; Mesobuthus.
REFERENCE 1 (bases 1 to 523)
AUTHORS Gantenbein,B., Thomson,M., Rosie,A., Parkinson,J. and Blaxter,M.
TITLE Gene discovery and phylogenetics of chelicerate arthropods using
expressed sequence tags
JOURNAL Unpublished (2002)
COMMENT Contact: Gantenbein-Ritter B
Institute of Cell, Animal and Population Biology
University of Edinburgh
West Mains Rd, Edinburgh, UK
Email: B.Gantenbein@ed.ac.uk
The library was prepared by Benjamin Gantenbein-Ritter and Marian
Thomson, University of Edinburgh, from the Metasoma + Telson of an
adult female. Sequencing was performed in Mark Blaxters lab in
Edinburgh.
PCR Primers
FORWARD: M13F
BACKWARD: pDNRLib_M13R
Plate: 30 row: D column: 09
Seq primer: M13F
High quality sequence stop: 370.
FEATURES
source
1..523
Location/Qualifiers
/organism="Mesobuthus gibbosus"
/mol_type="mRNA"
/strain="Farga"
/db_xref="taxon:123226"
/clone="Mg_AFT_30D09"
/sex="female"
/tissue_type="Metasoma + Telson"
/dev_stage="adult"
/clone_lib="Mesobuthus gibbosus adult female tail"
/notes="The Mesobuthus gibbosus EST library was prepared by
Benjamin Gantenbein and Marian Thomson, University of
Edinburgh. cDNA from the metasoma and telson of an adult
female was cloned into the vector pDNRLib. ESTs were
amplified from each clone by PCR, using primers pDNRLib.M13F
and pDNRLib.M13R. The products were cleaned of excess
nucleotides and phosphates, by treatment with SAP/ExoI,
and then sequenced using pDNRLib.seq primer."

ORIGIN
Query Match 75.2%; Score 18.8; DB 6; Length 523;
Best Local Similarity 90.9%; Pred. No. 1.6e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACAATGACCAACTGGGCTG 24
||||| ||||| ||||| |||||
Db 43 AAAAATAATGTCCAACTGGGCTG 22

RESULT 50
AG914898 687 bp DNA linear GSS 01-FEB-2005
LOCUS AG914898
DEFINITION Drosophila auraria DNA, clone: DABL-011L05.R.f.a, genomic survey
sequence.
ACCESSION AG914898
VERSION AG914898.1 GI:58436254

```

KEYWORDS  
SOURCE Drosophila auraria  
ORGANISM Drosophila auraria  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 Hattori, M., Toyoda, A., Murakami, K., Kuroki, Y., Fujiyama, A.,  
Toshio, T. K. and Sakaki, Y.  
BAC end sequences of Library DAB1  
Unpublished  
2 (bases 1 to 687)  
Hattori, M.  
Direct Submission  
Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the BAC library DAB1  
For BAC library availability, please contact Masa-Toshi Yamamoto  
(yamamoto@kit.jp).  
Submitted (30-11-2004) by Masahira Hattori,  
RIKEN, Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, Tel: 81-45-503-9111,  
Fax: 81-45-503-9170)  
This work was done in collaboration with Yamamoto, M.-T. Drosophila  
Genetic Resource Center  
Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan  
Tel: 81-75-873-2660 FAX: 81-75-861-0881  
PRIMERS  
Sequencing : R  
LIBRARY  
Vector : pKS150  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .687  
/organism="Drosophila auraria"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:47315"  
/clone="DAB1-011L05.R.fa"  
/clone\_lib="DAB1 Drosophila BAC library"  
ORIGIN  
Query Match 75.2%; Score 18.8; DB 10; Length 687;  
Best Local Similarity 90.9%; Pred. No. 1.6e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAACAATGACCAACTGGGCT 23  
|||||  
Db 273 AAAAAACATTGACCAACTGGTCT 294  
Search completed: February 3, 2006, 22:01:55  
Job time : 2961.67 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-5

Perfect score: 25

Sequence: 1 aaaaacaatgacaaactgggctgt 25

Scoring table: IDENTIFY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfileseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22.4	89.6	601	3	US-09-949-016-128244
2	20.8	83.2	601	3	US-09-949-016-128243
C 3	20.8	83.2	11994	3	US-09-949-016-15373
C 4	18.8	75.2	26843	3	US-09-949-016-17208
5	18.6	74.4	87523	3	US-09-949-016-12670
6	18.6	74.4	87523	3	US-09-949-016-15047
7	18.6	74.4	87523	3	US-09-949-016-15048
8	18.6	74.4	87523	3	US-09-949-016-15049
9	18.6	74.4	87869	3	US-09-949-016-11744
10	18.6	74.4	87869	3	US-09-949-016-15044
11	18.6	74.4	87869	3	US-09-949-016-15045
12	18.6	74.4	87869	3	US-09-949-016-15046
C 13	18.4	73.6	601	3	US-09-949-016-191078
C 14	17.6	70.4	601	3	US-09-949-016-40234
C 15	17.6	70.4	601	3	US-09-949-016-51844
C 16	17.6	70.4	601	3	US-09-949-016-139379
C 17	17.6	70.4	601	3	US-09-949-016-154598
C 18	17.6	70.4	3256	3	US-10-104-047-474
C 19	17.6	70.4	13335	3	US-09-949-016-13162
C 20	17.6	70.4	80269	3	US-09-949-016-15681
21	17.6	70.4	132456	3	US-09-949-016-13750
22	17.6	70.4	146095	3	US-09-949-016-12872
23	17.6	70.4	146104	3	US-09-949-016-13239
24	17.6	70.4	275110	3	US-09-949-016-12706

25	17.6	70.4	275110	3	US-09-949-016-16070	Sequence 16070, A
C 26	17.2	68.8	601	3	US-09-949-016-129135	Sequence 129135, A
27	17.2	68.8	3381	3	US-09-009-119-1	Sequence 1, Appli
28	17.2	68.8	3381	3	US-09-371-507-1	Sequence 1, Appli
29	17.2	68.8	3383	6	PCT-US95-09098-1	Sequence 1, Appli
30	17.2	68.8	14079	3	US-09-949-016-11993	Sequence 11993, A
C 31	17.2	68.8	98864	3	US-09-949-016-15403	Sequence 15403, A
C 32	17	68.0	500	3	US-10-002-623-908	Sequence 908, App
C 33	17	68.0	601	3	US-09-949-016-133748	Sequence 133748, A
C 34	17	68.0	601	3	US-09-949-016-133772	Sequence 133772, A
35	17	68.0	601	3	US-09-949-016-151005	Sequence 151005, A
C 36	17	68.0	601	3	US-09-949-016-169748	Sequence 169748, A
C 37	17	68.0	601	3	US-09-949-016-137059	Sequence 137059, A
C 38	17	68.0	700	3	US-09-735-271-884	Sequence 884, App
C 39	17	68.0	8888	3	US-09-949-016-12308	Sequence 12308, A
C 40	17	68.0	8889	3	US-09-949-016-15760	Sequence 15760, A
41	17	68.0	15585	3	US-09-949-016-11927	Sequence 11927, A
42	17	68.0	15585	3	US-09-949-016-15627	Sequence 15627, A
43	17	68.0	36023	3	US-09-949-016-15577	Sequence 15577, A
C 44	17	68.0	45842	3	US-09-949-016-12550	Sequence 12550, A
C 45	17	68.0	45842	3	US-09-949-016-17327	Sequence 17327, A
46	17	68.0	56241	3	US-09-949-016-15174	Sequence 15174, A
47	17	68.0	56241	3	US-09-949-016-15175	Sequence 15175, A
C 48	17	68.0	75295	3	US-09-949-002-575	Sequence 575, App
C 49	17	68.0	75296	3	US-09-949-002-799	Sequence 799, App
50	17	68.0	99748	3	US-09-949-016-11990	Sequence 11990, A
51	17	68.0	99748	3	US-09-949-016-16518	Sequence 16518, A
52	17	68.0	116652	3	US-09-949-016-13413	Sequence 13413, A
53	17	68.0	183770	3	US-09-949-016-15494	Sequence 15494, A
54	17	68.0	239527	3	US-09-949-016-15980	Sequence 15980, A
C 55	17	68.0	323820	3	US-09-949-016-14139	Sequence 14139, A
C 56	17	68.0	421491	3	US-09-949-016-12805	Sequence 12805, A
57	17	68.0	421494	3	US-09-949-016-14060	Sequence 14060, A
58	16.8	67.2	652	3	US-09-270-767-4915	Sequence 4915, Ap
59	16.8	67.2	652	3	US-09-270-767-20197	Sequence 20197, A
60	16.8	67.2	1764	3	US-09-579-383-2	Sequence 2, Appli
C 61	16.8	67.2	2500	3	US-09-579-383-5	Sequence 5, Appli
C 62	16.8	67.2	193555	3	US-09-949-016-15553	Sequence 15553, A
C 63	16.8	67.2	193555	3	US-09-949-016-15554	Sequence 15554, A
C 64	16.8	67.2	193555	3	US-09-949-016-15555	Sequence 15555, A
C 65	16.6	66.4	396	3	US-09-583-110-234	Sequence 234, App
C 66	16.6	66.4	601	3	US-09-949-016-68220	Sequence 68220, A
C 67	16.6	66.4	601	3	US-09-949-016-68221	Sequence 68221, A
C 68	16.6	66.4	601	3	US-09-949-016-68222	Sequence 68222, A
C 69	16.6	66.4	601	3	US-09-949-016-76457	Sequence 76457, A
70	16.6	66.4	601	3	US-09-949-016-85312	Sequence 85312, A
71	16.6	66.4	1917	3	US-09-270-767-15042	Sequence 15042, A
72	16.6	66.4	9279	3	US-09-487-558B-171	Sequence 171, App
C 73	16.6	66.4	10764	3	US-09-949-016-11872	Sequence 11872, A
C 74	16.6	66.4	10765	3	US-09-949-016-14204	Sequence 14204, A
C 75	16.6	66.4	11378	3	US-08-961-527-210	Sequence 210, App
C 76	16.6	66.4	31861	3	US-09-949-016-12803	Sequence 12803, A
C 77	16.6	66.4	31861	3	US-09-949-016-13967	Sequence 13967, A
C 78	16.6	66.4	37288	3	US-09-949-016-14593	Sequence 14593, A
C 79	16.6	66.4	37288	3	US-09-949-016-14594	Sequence 14594, A
C 80	16.6	66.4	71863	3	US-09-949-016-15112	Sequence 15112, A
C 81	16.6	66.4	83851	3	US-09-949-016-13847	Sequence 13847, A
C 82	16.6	66.4	88557	3	US-09-949-016-17028	Sequence 17028, A
C 83	16.6	66.4	94830	3	US-09-949-016-12414	Sequence 12414, A
C 84	16.6	66.4	94847	3	US-09-949-016-16336	Sequence 16336, A
C 85	16.6	66.4	115954	3	US-09-949-016-12298	Sequence 12298, A
C 86	16.6	66.4	119594	3	US-09-949-016-12080	Sequence 12080, A
C 87	16.6	66.4	119594	3	US-09-949-016-15952	Sequence 15952, A
C 88	16.6	66.4	119601	3	US-09-949-016-14711	Sequence 14711, A
C 89	16.6	66.4	156334	3	US-09-949-016-13749	Sequence 13749, A
C 90	16.6	66.4	194933	3	US-09-949-016-14172	Sequence 14172, A
C 91	16.6	66.4	200663	3	US-09-949-016-12569	Sequence 12569, A
C 92	16.6	66.4	213456	3	US-09-820-007-3	Sequence 3, Appli
C 93	16.6	66.4	294816	3	US-09-949-016-15974	Sequence 15974, A
94	16.6	66.4	330416	3	US-09-949-016-16923	Sequence 16923, A
95	16.4	65.6	1490	3	US-10-200-012-34	Sequence 34, Appli
96	16.4	65.6	1652	3	US-09-774-528-17	Sequence 17, Appli
97	16.4	65.6	1652	3	US-10-120-988-17	Sequence 17, Appli





LENGTH: 11994  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15373

Query Match 83.2%; Score 20.8; DB 3; Length 11994;  
Best Local Similarity 91.7%; Pred. No. 9.7;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24  
DB 8548 AAAAAACAATGTCACCACTGGTCTG 8525

## RESULT 4

US-09-949-016-17208/c  
; Sequence 17208, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17208  
; LENGTH: 26843  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17208

Query Match 75.2%; Score 18.8; DB 3; Length 26843;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCT 23  
DB 13095 AAAAAACAATGACCACTGGGCT 13074

## RESULT 5

US-09-949-016-12670  
; Sequence 12670, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12670  
; LENGTH: 87523  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12670

Query Match 74.4%; Score 18.6; DB 3; Length 87523;  
Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25  
DB 6500 AAAAAACAATCACCACCAATGGGATGT 6524

## RESULT 6

US-09-949-016-15047  
; Sequence 15047, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15047  
; LENGTH: 87523  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15047

Query Match 74.4%; Score 18.6; DB 3; Length 87523;  
Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25  
DB 6500 AAAAAACAATCACCACCAATGGGATGT 6524

## RESULT 7

US-09-949-016-15048  
; Sequence 15048, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15048  
; LENGTH: 87523  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15048

Query Match 74.4%; Score 18.6; DB 3; Length 87523;  
Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25

```
Db      6500 AAAAAAATACCAAAATGGGATGT 6524
||||| ||| ||||| ||||| ||||| |||||
US-09-949-016-15049
; Sequence 15049, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15049
; LENGTH: 87523
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15049

Query Match      74.4%; Score 18.6; DB 3; Length 87523;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAACAATGACCAACTGGGCTGT 25
||||| ||| ||||| ||||| ||||| |||||
Db      6500 AAAAAAATACCAAAATGGGATGT 6524

RESULT 9
US-09-949-016-11744
; Sequence 11744, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11744
; LENGTH: 87869
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11744

Query Match      74.4%; Score 18.6; DB 3; Length 87869;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAACAATGACCAACTGGGCTGT 25
||||| ||| ||||| ||||| ||||| |||||
Db      6846 AAAAAAATACCAAAATGGGATGT 6870

RESULT 10
US-09-949-016-15046
; Sequence 15046, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
US-09-949-016-15044
; Sequence 15044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15044
; LENGTH: 87869
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15044

Query Match      74.4%; Score 18.6; DB 3; Length 87869;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAACAATGACCAACTGGGCTGT 25
||||| ||| ||||| ||||| ||||| |||||
Db      6846 AAAAAAATACCAAAATGGGATGT 6870

RESULT 11
US-09-949-016-15045
; Sequence 15045, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15045
; LENGTH: 87869
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15045

Query Match      74.4%; Score 18.6; DB 3; Length 87869;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAACAATGACCAACTGGGCTGT 25
||||| ||| ||||| ||||| ||||| |||||
Db      6846 AAAAAAATACCAAAATGGGATGT 6870

RESULT 12
US-09-949-016-15046
; Sequence 15046, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15046
; LENGTH: 87869
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15046

Query Match          74.4%; Score 18.6; DB 3; Length 87869;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 6846 AAAAAAATACCAAAATGGGATGT 6870
||||| ||| ||||| ||||| |||||

RESULT 13
US-09-949-016-191078/c
; Sequence 191078, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191078
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-191078

Query Match          73.6%; Score 18.4; DB 3; Length 601;
Best Local Similarity 86.4%; Pred. No. 78;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCT 23
Db 306 AAAAAAAGACCACTGGGCT 285
||||| ||| ||||| ||||| |||||

RESULT 14
US-09-949-016-40234/c
; Sequence 40234, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40234
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40234

Query Match          70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
Db 253 AAAAAACAATGACCAACCGGCTG 230
||||| ||| ||||| ||||| |||||

RESULT 15
US-09-949-016-51844/c
; Sequence 51844, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51844
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-51844

Query Match          70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24
Db 253 AAAAAACAATGACCAACCGGCTG 230
||||| ||| ||||| ||||| |||||

RESULT 16
US-09-949-016-139379
; Sequence 139379, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 139379
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-139379

Query Match          70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
    ||||| ||||| ||||| ||||| |||||
Db 560 AAAACAATGAACAGCTAAGCTGT 583

RESULT 17
US-09-949-016-154598/c
; Sequence 154598, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154598
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-154598

Query Match          70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
    ||||| ||||| ||||| ||||| |||||
Db 117 AAAACAAGACAACTGGATTGT 94

RESULT 18
US-10-104-047-474
; Sequence 474, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 474
; LENGTH: 3256
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-474

Query Match          70.4%; Score 17.6; DB 3; Length 3256;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
    ||||| ||||| ||||| ||||| |||||
Db 1225 AAAACAATGATCGCCAGGCTGT 1248

Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
    ||||| ||||| ||||| ||||| |||||
Db 1225 AAAACAATGATCGCCAGGCTGT 1248

RESULT 19
US-09-949-016-13162/c
; Sequence 13162, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13162
; LENGTH: 13335
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13162

Query Match          70.4%; Score 17.6; DB 3; Length 13335;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAACAATGACCAACTGGGCTG 24
    ||||| ||||| ||||| ||||| |||||
Db 3954 AAAAAAAGCAACTGGGCTG 3931

RESULT 20
US-09-949-016-15681
; Sequence 15681, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15681
; LENGTH: 80269
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15681

Query Match          70.4%; Score 17.6; DB 3; Length 80269;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
    ||||| ||||| ||||| ||||| |||||
```

Db 73866 AAAAAACAATGACACGCTAAGCTGT 73889

## RESULT 21

US-09-949-016-13750  
; Sequence 13750, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13750  
; LENGTH: 132456  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(132456)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13750

Query Match 70.4%; Score 17.6; DB 3; Length 132456;  
Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 47727 AACAACAATGACCAACTGGGCTGT 47750

## RESULT 22

US-09-949-016-12872  
; Sequence 12872, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12872  
; LENGTH: 146095  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12872

Query Match 70.4%; Score 17.6; DB 3; Length 146095;  
Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24  
|||||  
Db 140538 AAAAAACAATGACCAACTGGGCTG 140561

## RESULT 23

US-09-949-016-13239  
; Sequence 13239, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13239  
; LENGTH: 146104  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13239

Query Match 70.4%; Score 17.6; DB 3; Length 146104;  
Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTG 24  
|||||  
Db 140538 AAAAAACAATGACCAACTGGGCTG 140561

## RESULT 24

US-09-949-016-12706  
; Sequence 12706, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12706  
; LENGTH: 275110  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12706

Query Match 70.4%; Score 17.6; DB 3; Length 275110;  
Best Local Similarity 83.3%; Pred. No. 3.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGCTGT 25  
|||||  
Db 27732 AAAAAACAAGACAAACTGGATTGT 27755

## RESULT 25

US-09-949-016-16070  
; Sequence 16070, Application US/09949016

```

; Patent No. 6812339
;
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;
; FILE REFERENCE: CU001307
;
; CURRENT APPLICATION NUMBER: US/09/949,016
;
; CURRENT FILING DATE: 2000-04-14
;
; PRIOR APPLICATION NUMBER: 60/241,755
;
; PRIOR FILING DATE: 2000-10-20
;
; PRIOR APPLICATION NUMBER: 60/237,768
;
; PRIOR FILING DATE: 2000-10-03
;
; PRIOR APPLICATION NUMBER: 60/231,498
;
; PRIOR FILING DATE: 2000-09-08
;
; NUMBER OF SEQ ID NOS: 207012
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 16070
;
; LENGTH: 275110
;
; TYPE: DNA
;
; ORGANISM: Human
;
; US-09-949-016-16070
;

```

```

; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
; FILE REFERENCE: substitute sequence listing
; CURRENT APPLICATION NUMBER: US/09/009,119
; CURRENT FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3381
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-09-009-119-1

Query Match      68.8%; Score 17.2; DB 3; Length 3381;
Best Local Similarity 86.4%; Pred.No.3.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      4 AAACAATGCACCACTGGGCTGT 25
        ||||||| ||||| |||||
Db      340 AAACAATGCACACGGGCTGT 361

RESULT 28
US-09-371-507-1
; Sequence 1, Application US/09371507
; Patent No. 6346656
; GENERAL INFORMATION:
; APPLICANT: SATO, RYO
; APPLICANT: BOYNTON, John E.
; APPLICANT: GILLHAM, Nicholas W.
; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
; FILE REFERENCE: substitute sequence listing
; CURRENT APPLICATION NUMBER: US/09/371,507
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 09/009,119
; PRIOR FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3381
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-09-371-507-1

Query Match      68.8%; Score 17.2; DB 3; Length 3381;
Best Local Similarity 86.4%; Pred.No.3.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      4 AAACAATGCACCACTGGGCTGT 25
        ||||||| ||||||| |||||
Db      340 AAACAATGCACACGGGCTGT 361

RESULT 29
PCT-US95-09098-1
; Sequence 1, Application PC/TUS9509098
; GENERAL INFORMATION:
; APPLICANT: Sato, Ryo
; APPLICANT: Boynton, John E.
; APPLICANT: Gillham, Nicholas W.
; APPLICANT: Harris, Elizabeth H.
; TITLE OF INVENTION: Porphyrin-Accumulating Type Herbicide
; TITLE OF INVENTION: Resistance Gene
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:

```





```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33748
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-33748

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 193 AATAACAATGATGAATGGTCTGT 169

RESULT 34
US-09-949-016-133372
; Sequence 133372, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133372
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-133372

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 181 AAAAAACAATGACCAACTGGGCTGT 205

RESULT 35
US-09-949-016-151005
; Sequence 151005, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151005
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151005

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 576 AGAAAAGCATGACCCAGTGGGCTGT 600

RESULT 36
US-09-949-016-169748/c
; Sequence 169748, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169748
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-169748

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 311 AAAAAAATAACCAACTGGGCTAT 287

RESULT 37
US-09-949-016-197059/c
; Sequence 197059, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

Query Match 68.0%; Score 17; DB 3; Length 8888;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 17: Conservative 0; Mismatches 0; Indels

```
Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20: Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

Qy 1 AAAAAACAATGACCAAC 17  
Dp 6956 AAAAAACAATGACCAAC 6940

```

RESULT 40
US-09-949-016-15760/c
; Sequence 15760, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15760
; LENGTH: 8889
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15760

```

Query Match 68.0%; Score 17; DB 3; Length 8889;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 17; Conservative 0; Mismatches 0 Indels

Qy 1 AAAAAACAATGACCAAC 17  
db 6956 AAAAAACAATGACCAAC 6940

```

RESULT 41
US-09-949-016-11927
; Sequence 11927, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN
; FILE REFERENCE: WITH HUMAN DISEASE, I
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11927
; LENGTH: 15585
; TYPE: DNA

```

```

: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASES, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11927
: LENGTH: 15585
: TYPE: DNA

```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(15585)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11927

Query Match      68.0%; Score 17; DB 3; Length 15585;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
    ||||| || ||||| || |||||
Db 10984 AAAAAACAGTGTCACAGGGGACTGT 11008

RESULT 42
US-09-949-016-15627
; Sequence 15627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15627
; LENGTH: 15585
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(15585)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15627

Query Match      68.0%; Score 17; DB 3; Length 15585;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
    ||||| || ||||| || |||||
Db 10984 AAAAAACAGTGTCACAGGGGACTGT 11008

RESULT 43
US-09-949-016-15577
; Sequence 15577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15577
; LENGTH: 15577
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(15577)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15577

Query Match      68.0%; Score 17; DB 3; Length 15585;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
    ||||| || ||||| || |||||
Db 10984 AAAAAACAGTGTCACAGGGGACTGT 11008

RESULT 44
US-09-949-016-12550/c
; Sequence 12550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12550
; LENGTH: 45842
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(45842)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12550

Query Match      68.0%; Score 17; DB 3; Length 45842;
Best Local Similarity 80.0%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
    ||||| || ||||| || |||||
Db 41349 AATAACACATGATGAAATGGTCTGT 41325

RESULT 45
US-09-949-016-17327/c
; Sequence 17327, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17327
```

```
; LENGTH: 45842
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-17327

Query Match      68.0%; Score 17; DB 3; Length 45842;
Best Local Similarity 80.0%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 41349 AATAACAATGATGAAATGGTCTGT 41325

RESULT 46
US-09-949-016-15174
; Sequence 15174, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15174
; LENGTH: 56241
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(56241)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15174

Query Match      68.0%; Score 17; DB 3; Length 56241;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 41244 AAAAAATCATGACCAACTGGACTCT 41268

RESULT 47
US-09-949-016-15175
; Sequence 15175, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15175
; LENGTH: 56241
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(56241)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15175
```

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(56241)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15175

Query Match      68.0%; Score 17; DB 3; Length 56241;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 41244 AAAAAATCATGACCAACTGGACTCT 41268

RESULT 48
US-09-949-002-575/c
; Sequence 575, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575
; LENGTH: 75295
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(75295)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-575

Query Match      68.0%; Score 17; DB 3; Length 75295;
Best Local Similarity 80.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 1516 AATAGCAATGATCAAAATGGGCTCT 1492

RESULT 49
US-09-949-002-799/c
; Sequence 799, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 799
; LENGTH: 75296
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (1) _ (75296)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-799

Query Match      68.0%; Score 17; DB 3; Length 75296;
Best Local Similarity 80.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||| ||| ||| ||| ||| ||| |||
Db 1516 AATAGCAATGATCAAAATGGGCTCT 1492

RESULT 50
US-09-949-016-11990
; Sequence 11990, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 11990
; LENGTH: 99748
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11990

Query Match      68.0%; Score 17; DB 3; Length 99748;
Best Local Similarity 80.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||| ||| ||| ||| ||| ||| |||
Db 60414 AAAAAAACAACCAACTGGGCTAT 60438

Search completed: February 3, 2006, 16:32:05
Job time : 88.6667 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-5

Perfect score: 25

Sequence: 1 aaaaacaatgacaaactgggctgt 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA\_Main:

- 1: /cgn2\_6/ptodata/1/pubpna/us07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/us08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/us09\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/us09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/us10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/us10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/us10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/us10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/us10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/us11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	8	US-10-719-900-5
2	23.4	93.6	25	8	US-10-719-900-6
3	20.8	83.2	598	4	US-09-925-065A-832538
4	20.2	80.8	553	4	US-09-925-065A-334942
5	20.2	80.8	553	4	US-09-925-065A-334943
6	20.2	80.8	585	4	US-09-925-065A-908094
7	19.4	77.6	2316	7	US-10-437-963-67029
8	19.2	76.8	460	3	US-10-437-963-50993
9	19.2	76.8	600	3	US-09-864-761-13549
10	19.2	76.8	1182	7	US-10-437-963-55485
11	18.8	75.2	6596	10	US-11-097-143-9349
12	18.6	74.4	171843	9	US-10-981-277-44
13	18.6	74.4	290547	7	US-10-367-094-77
14	18.4	73.6	308	3	US-09-954-531-4
15	18.4	73.6	308	9	US-10-843-641A-1071
16	18.4	73.6	327	7	US-10-242-535A-12176
17	18.4	73.6	327	7	US-10-085-783A-12176
18	18.4	73.6	414	8	US-10-425-115-49124
19	18.4	73.6	471	7	US-10-242-535A-43265
20	18.4	73.6	471	7	US-10-085-783A-43265
21	18.4	73.6	2146	8	US-10-723-860-4899
22	18.4	73.6	2178	8	US-10-425-115-183781
23	18.2	72.8	584	4	US-09-925-065A-187435

24	18.2	72.8	584	4	US-09-925-065A-187436	Sequence 187436,
25	18.2	72.8	594	4	US-09-925-065A-827160	Sequence 827160,
26	18.2	72.8	607	4	US-09-925-065A-584904	Sequence 584904,
27	18.2	72.8	613	4	US-09-925-065A-309641	Sequence 309641,
28	18.2	72.8	650	7	US-10-021-323-17315	Sequence 17315, A
29	18.2	72.8	1003	7	US-10-282-122A-36969	Sequence 36969, A
30	18.2	72.8	1047	7	US-10-282-122A-39716	Sequence 39716, A
31	18.2	72.8	1286	4	US-09-925-065A-676311	Sequence 676311,
32	18.2	72.8	1286	4	US-09-925-065A-676312	Sequence 676312,
33	18.2	72.8	1286	4	US-09-925-065A-676313	Sequence 676313,
34	18.2	72.8	1286	4	US-09-925-065A-676314	Sequence 676314,
35	18.2	72.8	2319	7	US-10-425-114-3974	Sequence 3974, Ap
36	18.2	72.8	2526	8	US-10-425-115-59884	Sequence 59884, A
37	18.2	72.8	54810	8	US-10-417-375-91	Sequence 91, Appl
38	18.2	72.8	66479	5	US-10-041-856-1	Sequence 1, Appl
39	18.2	72.8	66479	10	US-11-073-203-1	Sequence 1, Appl
40	18.2	72.8	78878	8	US-10-719-993-6863	Sequence 6863, Ap
41	18.2	72.8	608916	9	US-10-461-862-1	Sequence 1, Appl
42	18	72.0	1522	3	US-09-938-842A-3818	Sequence 3818, Ap
43	18	72.0	1522	3	US-09-938-842A-3818	Sequence 3818, Ap
44	17.8	71.2	201	8	US-10-741-600-33597	Sequence 33597, A
45	17.8	71.2	379	8	US-10-425-115-171078	Sequence 171078,
46	17.8	71.2	415	8	US-10-425-115-47284	Sequence 47284, A
47	17.8	71.2	612	4	US-09-925-065A-926511	Sequence 926511,
48	17.8	71.2	2570	7	US-10-437-963-41987	Sequence 41987, A
49	17.8	71.2	7600	8	US-10-723-860-6853	Sequence 6853, Ap
50	17.8	71.2	99998	6	US-10-085-117-34	Sequence 34, Appl
51	17.8	71.2	183334	8	US-10-741-600-17646	Sequence 17646, A
52	17.6	70.4	232	8	US-10-425-115-51038	Sequence 51038, A
53	17.6	70.4	374	4	US-09-925-065A-250403	Sequence 250403,
54	17.6	70.4	397	3	US-09-960-352-835	Sequence 835, App
55	17.6	70.4	459	7	US-10-425-114-21919	Sequence 21919, A
56	17.6	70.4	473	5	US-10-425-114-21977	Sequence 21977, A
57	17.6	70.4	516	7	US-10-027-632-139616	Sequence 139616,
58	17.6	70.4	516	5	US-10-027-632-139617	Sequence 139617,
59	17.6	70.4	516	6	US-10-027-632-139616	Sequence 139616,
60	17.6	70.4	534	4	US-09-925-065A-549550	Sequence 549550,
61	17.6	70.4	555	3	US-09-764-868-534	Sequence 534, App
62	17.6	70.4	568	9	US-10-779-543-18513	Sequence 18513, A
63	17.6	70.4	570	4	US-09-925-065A-465338	Sequence 465338,
64	17.6	70.4	570	4	US-09-925-065A-465339	Sequence 465339,
65	17.6	70.4	570	4	US-09-925-065A-783513	Sequence 783513,
66	17.6	70.4	577	4	US-09-925-065A-116999	Sequence 116999,
67	17.6	70.4	581	4	US-09-925-065A-190858	Sequence 190858,
68	17.6	70.4	587	4	US-09-925-065A-203935	Sequence 203935,
69	17.6	70.4	587	5	US-10-027-632-203935	Sequence 203935,
70	17.6	70.4	587	5	US-10-027-632-203936	Sequence 203936,
71	17.6	70.4	587	6	US-10-027-632-203935	Sequence 203935,
72	17.6	70.4	587	6	US-10-027-632-203936	Sequence 203936,
73	17.6	70.4	592	8	US-10-425-115-83254	Sequence 83254, A
74	17.6	70.4	597	4	US-09-925-065A-866728	Sequence 866728,
75	17.6	70.4	600	9	US-10-972-079-40005	Sequence 40005, A
76	17.6	70.4	614	5	US-10-027-632-291895	Sequence 291895,
77	17.6	70.4	614	6	US-10-027-632-291895	Sequence 291895,
78	17.6	70.4	626	4	US-09-925-065A-544804	Sequence 544804,
79	17.6	70.4	628	4	US-09-925-065A-927904	Sequence 927904,
80	17.6	70.4	635	8	US-10-425-115-151662	Sequence 151662,
81	17.6	70.4	669	5	US-10-106-698-643	Sequence 643, App
82	17.6	70.4	741	4	US-09-925-065A-955248	Sequence 955248,
83	17.6	70.4	743	4	US-09-925-065A-8859	Sequence 8859, Ap
84	17.6	70.4	797	9	US-10-779-543-4097	Sequence 4097, Ap
85	17.6	70.4	893	5	US-10-027-632-171163	Sequence 171163,
86	17.6	70.4	893	5	US-10-027-632-171163	Sequence 171163,
87	17.6	70.4	893	6	US-10-027-632-171163	Sequence 171163,
88	17.6	70.4	1071	4	US-09-925-065A-553293	Sequence 553293,
89	17.6	70.4	1193	7	US-10-425-114-22701	Sequence 22701, A
90	17.6	70.4	1197	4	US-09-925-065A-690699	Sequence 690699,
91	17.6	70.4	1570	7	US-10-433-802-33	Sequence 33, Appl
92	17.6	70.4	1670	7	US-10-425-115-34282	Sequence 34282, A
93	17.6	70.4	1714	8	US-09-764-868-161	Sequence 161, App
94	17.6	70.4	1921	3	US-10-437-963-50361	Sequence 50361, A
95	17.6	70.4	2691	7	US-10-161-493-2929	Sequence 29, Appl
96	17.6	70.4	2789	7	US-10-161-493-2929	Sequence 29, Appl

```

; SEQ ID NO 5
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-5
Query Match      100.0%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.16; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

OY 1 AAAAAACAATGACCAACTGGGCTGT 25
    |||||
Db 1 AAAAAACAATGACCAACTGGGCTGT 25

RESULT 2
US-10-719-900-6
; Sequence 6, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-6
Query Match      93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.85; Mismatches 1; Indels 0; Gaps 0;
Matches 24; Conservative 0;

OY 1 AAAAAACAATGACCAACTGGGCTGT 25
    |||||
Db 1 AAAAAACAATGACCAACTGGGCTGT 25

RESULT 3
US-09-925-065A-832538
; Sequence 832538, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 832538
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-832538
```

```

Sequence 27, Appl
Sequence 474, App
Sequence 21, Appl
Sequence 21, Appl
Sequence 1597, Ap
Sequence 1271, Ap
Sequence 46, Appl
Sequence 1100, Ap
Sequence 28, Appl
Sequence 17865, A
Sequence 3, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 11, Appl
Sequence 11, Appl
Sequence 17991, A
Sequence 97, Appl
Sequence 5729, Ap
Sequence 420, App
Sequence 386143,
Sequence 103, App
Sequence 470865,
Sequence 861934,
Sequence 162723,
Sequence 162724,
Sequence 162723,
Sequence 162724,
Sequence 55258, A
Sequence 55259, A
Sequence 55260, A
Sequence 55261, A
Sequence 2623, Ap
Sequence 37, Appl
Sequence 2623, Ap
Sequence 4333, Ap
Sequence 229810,
Sequence 392728,
Sequence 392729,
Sequence 141128,
Sequence 141129,
Sequence 134080,
Sequence 134081,
Sequence 134082,
Sequence 196559,
Sequence 196560,
Sequence 196559,
Sequence 196560,
Sequence 407, App
Sequence 58397, A
Sequence 58396, A
Sequence 30346, A
Sequence 1174, Ap
Sequence 382, App
```

ALIGNMENTS

```

RESULT 1
US-10-719-900-5
; Sequence 5, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```





```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67926C.1
US-10-437-963-67029

Query Match      77.6%; Score 19.4; DB 7; Length 2316;
Best Local Similarity 95.2%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACAATGACCAACTGGGCT 23
   ||||| ||||| ||||| ||||| |||||
DB 834 AAAATAATGACCAACTGGGCT 854

RESULT 8
US-10-437-963-50993/c
; Sequence 50993, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yihua
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50993
; LENGTH: 460
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53428C.1
US-10-437-963-50993

Query Match      76.8%; Score 19.2; DB 7; Length 460;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| ||||| |||||
DB 71 AAAAATGTGTGACCAACTGGGCTGT 48

RESULT 9
US-09-864-761-13549
; Sequence 13549, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13549
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009178.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-13549

Query Match      76.8%; Score 19.2; DB 3; Length 600;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| ||||| |||||
DB 370 AAAACCATGACCAACTGGTTGT 393

RESULT 10
US-10-437-963-55485
; Sequence 55485, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 55485
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_57489C.1
US-10-437-963-55485
```

```
Query Match 76.8%; Score 19.2; DB 7; Length 1182;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCTGT 25
DB 1104 AAAAAAGTGTGACCAACTGGGCTGT 1127

RESULT 11
US-11-097-143-9349/c
; Sequence 9349, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9349
; LENGTH: 6596
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-9349

Query Match 75.2%; Score 18.8; DB 10; Length 6596;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAATGACCAACTGGGCT 23
DB 6184 AAAAAAATGACCAACTGGACT 6163

RESULT 12
US-10-981-277-44
; Sequence 44, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 44
; LENGTH: 171843
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-981-277-44

Query Match 74.4%; Score 18.6; DB 9; Length 171843;
Best Local Similarity 84.0%; Pred. No. 6.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
DB 134483 AAAAAAATGTCCAACAGGGTTGT 134507

RESULT 13
US-10-367-094-77
; Sequence 77, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 290547
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(290547)
; OTHER INFORMATION: n = A,T,C or G
US-10-367-094-77

Query Match 74.4%; Score 18.6; DB 7; Length 290547;
Best Local Similarity 84.0%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCTGT 25
DB 26055 AAAAAAATATCAACAACACTGGGCTGT 26079

RESULT 14
US-09-954-531-4/c
; Sequence 4, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-4
```

Query Match 73.6%; Score 18.4; DB 3; Length 308;  
Best Local Similarity 95.0%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAACAATGACCAACTGG 20  
Db 21 AAAAAACAATGTCCAACCTGG 2  
  
RESULT 15  
US-10-843-641A-1071/c  
; Sequence 1071, Application US/10843641A  
; Publication No. US20050064454A1  
; GENERAL INFORMATION:  
; APPLICANT: Avalon Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; FILE REFERENCE: 689290-189  
; CURRENT APPLICATION NUMBER: US/10/843,641A  
; CURRENT FILING DATE: 2004-05-12  
; PRIOR APPLICATION NUMBER: US/09/873,367  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US/09/954,531  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,436  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/962,832  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/964,824  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/09/967,768  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US/09/968,007  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,347  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/09/969,708  
; PRIOR FILING DATE: 2001-10-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 847  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1071  
; LENGTH: 308  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-843-641A-1071

Query Match 73.6%; Score 18.4; DB 9; Length 308;  
Best Local Similarity 95.0%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGG 20  
Db 21 AAAAAACAATGTCCAACCTGG 2

RESULT 16  
US-10-242-535A-12176  
; Sequence 12176, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12176  
; LENGTH: 327  
; TYPE: DNA  
; ORGANISM: Human  
US-10-242-535A-12176

Query Match 73.6%; Score 18.4; DB 7; Length 327;  
Best Local Similarity 95.0%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGG 20  
Db 146 AAAAAACAATGTCCAACCTGG 165

RESULT 17  
US-10-085-783A-12176  
; Sequence 12176, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12176  
; LENGTH: 327  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-12176

Query Match 73.6%; Score 18.4; DB 7; Length 327;  
Best Local Similarity 95.0%; Pred. No. 2.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGG 20  
Db 146 AAAAAACAATGTCCAACCTGG 165

RESULT 18  
US-10-425-115-49124  
; Sequence 49124, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369126  
; SEQ ID NO 49124  
; LENGTH: 414  
; TYPE: DNA

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_144806C.1
US-10-425-115-49124

Query Match          73.6%; Score 18.4; DB 8; Length 414;
Best Local Similarity 95.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGG 20
   |||||
Db 38 AAAAAACAATGTCCAACCTGG 57

RESULT 19
US-10-242-535A-43265
; Sequence 43265, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43265
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-43265

Query Match          73.6%; Score 18.4; DB 7; Length 471;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGG 20
   |||||
Db 164 AAAAAACAATGTCCAACCTGG 183

RESULT 20
US-10-085-783A-43265
; Sequence 43265, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43265
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Human
```

```
US-10-085-783A-43265

Query Match          73.6%; Score 18.4; DB 7; Length 471;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGG 20
   |||||
Db 164 AAAAAACAATGTCCAACCTGG 183

RESULT 21
US-10-723-860-4899/c
; Sequence 4899, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nataeha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4899
; LENGTH: 2146
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-4899

Query Match          73.6%; Score 18.4; DB 8; Length 2146;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGG 20
   |||||
Db 679 AAAAAACAATGTCCAACCTGG 660

RESULT 22
US-10-425-115-183781/c
; Sequence 183781, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 183781
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2178)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_99191C.1
US-10-425-115-183781

Query Match          73.6%; Score 18.4; DB 8; Length 2178;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAACAATGACCAACTGG 20
|||||
Db 2003 AAAAAACAATACCAACTGG 1984

RESULT 23
US-09-925-065A-187435
; Sequence 187435, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187435
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-187435

Query Match 72.8%; Score 18.2; DB 4; Length 584;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAACAATGACCAACTGGCTGT 25
|||||
Db 492 AATACAATGACCAACTGGCTTT 514

RESULT 24
US-09-925-065A-187436
; Sequence 187436, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187436
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-187436

Query Match 72.8%; Score 18.2; DB 4; Length 584;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAACAATGACCAACTGGCTGT 25
|||||
Db 492 AATACAATGACCAACTGGCTTT 514

RESULT 25
US-09-925-065A-827160
; Sequence 827160, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 827160
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-827160

Query Match 72.8%; Score 18.2; DB 4; Length 594;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGCT 23
|||||
Db 21 AAAAAAATGACCAACTGGACT 43

RESULT 26
US-09-925-065A-584904
; Sequence 584904, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 584904
; LENGTH: 607
; TYPE: DNA
```





```
RESULT 30
US-10-282-122A-39716
; Sequence 39716, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39716
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Salmonella typhi
US-10-282-122A-39716

Query Match 72.8%; Score 18.2; DB 7; Length 1047;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAATGACCAACTGGGCTG 24
||||| ||||| ||||| ||||| |||||
Db 435 AAAAAACATTACCAGCTGGGCTG 457

RESULT 31
US-09-925-065A-676311/c
; Sequence 676311, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US/09/925,065A
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 676312
; LENGTH: 1286
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-676312

Query Match 72.8%; Score 18.2; DB 4; Length 1286;
Best Local Similarity 87.0%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCT 23
||||| ||||| ||||| ||||| |||||
Db 1217 AAAAAACCATGACCAACTGAGGT 1195

RESULT 32
US-09-925-065A-676312/c
; Sequence 676312, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 676312
; LENGTH: 1286
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-676312

Query Match 72.8%; Score 18.2; DB 4; Length 1286;
Best Local Similarity 87.0%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCT 23
||||| ||||| ||||| ||||| |||||
Db 1217 AAAAAACCATGACCAACTGAGGT 1195

RESULT 33
US-09-925-065A-676313/c
; Sequence 676313, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
```

APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 3974  
LENGTH: 2319  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700345691\_FLI  
US-10-425-114-3974

Query Match 72.8%; Score 18.2; DB 7; Length 2319;  
Best Local Similarity 87.0%; Pred. No. 4.7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23  
DB 2012 AAAAAACAATGACCAACTGGGCT 2034

## RESULT 36

US-10-425-115-59884  
Sequence 59884, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 59884  
LENGTH: 2526  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(2526)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_154612C.1  
US-10-425-115-59884

Query Match 72.8%; Score 18.2; DB 8; Length 2526;  
Best Local Similarity 87.0%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23  
DB 2012 AAAAAACAATGACCAACTGGGCT 2034

## RESULT 37

US-10-417-375-91/c  
Sequence 91, Application US/10417375  
Publication No. US20040219528A1  
GENERAL INFORMATION:  
APPLICANT: David W. Morris  
APPLICANT: Marc Malandro  
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
FILE REFERENCE: 529452001600  
CURRENT APPLICATION NUMBER: US/10/417,375  
CURRENT FILING DATE: 2003-04-15

PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 676313  
LENGTH: 1286  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-676313

Query Match 72.8%; Score 18.2; DB 4; Length 1286;  
Best Local Similarity 87.0%; Pred. No. 4.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23  
DB 1217 AAAAAACAATGACCAACTGGGT 1195

## RESULT 34

US-09-925-065A-676314/c  
Sequence 676314, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 676314  
LENGTH: 1286  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-676314

Query Match 72.8%; Score 18.2; DB 4; Length 1286;  
Best Local Similarity 87.0%; Pred. No. 4.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23  
DB 1217 AAAAAACAATGACCAACTGGGT 1195

## RESULT 35

US-10-425-114-3974  
Sequence 3974, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.



Db 432617 AACAGCAATGACCAACTGGCTG 432639

## RESULT 42

US-09-938-842A-3818/c  
; Sequence 3818, Application US/09938842A  
; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; SAME, AND METHODS OF USE

; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 3818

; LENGTH: 1522

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-3818

Query Match 72.0%; Score 18; DB 3; Length 1522;

Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACT 18

|||||

Db 1386 AAAAAACAATGACCAACT 1369

## RESULT 43

US-09-938-842A-3818/c

; Sequence 3818, Application US/09938842A

; Publication No. US20040009476A9

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; SAME, AND METHODS OF USE

; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 3818

; LENGTH: 1522

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-3818

Query Match 72.0%; Score 18; DB 3; Length 1522;

Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACT 18

|||||

Db 1386 AAAAAACAATGACCAACT 1369

## RESULT 44

US-10-741-600-33597/c

; Sequence 33597, Application US/10741600

; Publication No. US20050026169A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001499

; CURRENT APPLICATION NUMBER: US/10/741,600

; CURRENT FILING DATE: 2003-12-22

; NUMBER OF SEQ ID NOS: 73997

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 33597

; LENGTH: 201

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-741-600-33597

Query Match 71.2%; Score 17.8; DB 8; Length 201;

Best Local Similarity 82.6%; Pred. No. 4.6e+02;

Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23

|||||

Db 112 AAAAAACAACARCAAACTGGGCT 90

## RESULT 45

US-10-425-115-171078/c

; Sequence 171078, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

PLANTS

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 171078

; LENGTH: 379

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577\_87601C.1

US-10-425-115-171078

Query Match 71.2%; Score 17.8; DB 8; Length 379;

Best Local Similarity 90.5%; Pred. No. 5.2e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACAAATGACCAACTGGGCTGT 25

|||||

Db 256 AACAAATGACCAACTGGGCGAGT 236

## RESULT 46

US-10-425-115-47284

; Sequence 47284, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

PLANTS

```
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 47284
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_143129C.1
US-10-425-115-47284

Query Match          71.2%; Score 17.8; DB 8; Length 415;
Best Local Similarity 90.5%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGC 22
   ||||| ||||| ||||| |||||
Db 114 AAAAAACAATTAACAACCTGGGC 134

RESULT 47
US-09-925-065A-926511/c
; Sequence 926511, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 926511
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-926511

Query Match          71.2%; Score 17.8; DB 4; Length 612;
Best Local Similarity 82.6%; Pred. No. 5.7e+02;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCT 23
   ||||| ||||| : ||||| |||||
Db 186 AAAAAACAACARCAAACTGGGCT 164

RESULT 48
US-10-437-963-41987/c
; Sequence 41987, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 41987
; LENGTH: 2570
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_45280C.1
US-10-437-963-41987

Query Match          71.2%; Score 17.8; DB 7; Length 2570;
Best Local Similarity 90.5%; Pred. No. 7.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGG 21
   ||||| ||||| ||||| |||||
Db 1093 AAAAAATAATGCCCACTGCG 1073

RESULT 49
US-10-723-860-6853/c
; Sequence 6853, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6853
; LENGTH: 7600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2002)..(2034)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-6853

Query Match          71.2%; Score 17.8; DB 8; Length 7600;
Best Local Similarity 90.5%; Pred. No. 8.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACAAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| |||||
Db 4359 AACAAAGGACCCCACTGGGCTGT 4339

RESULT 50
US-10-085-117-34/c
; Sequence 34, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
```

Mon Feb 6 12:23:20 2006

```
; SEQ ID NO 34
; LENGTH: 99998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(99998)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-34

Query Match      71.2%; Score 17.8; DB 6; Length 99998;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 AAAAACAATGACCACTGGGC 22
      ||||| ||||| ||||| |||||
Db      98973 AAAAACAACCACTGGGC 98953

Search completed: February 3, 2006, 15:43:34
Job time : 371.556 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-5

Perfect score: 25

Sequence: 1 aaaaacaatgacaaactgggctgt 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 41903697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	17.8	71.2	100000	8	US-11-124-368A-2898 Sequence 2898, Ap
C 2	17.8	71.2	186442	8	US-11-121-086-104 Sequence 104, Ap
C 3	17.6	70.4	201	7	US-10-995-561-82214 Sequence 82214, A
C 4	17.6	70.4	195235	7	US-10-995-561-13495 Sequence 13495, A
C 5	17.6	70.4	403278	7	US-10-995-561-13421 Sequence 13421, A
C 6	17.4	69.6	83528	7	US-10-995-561-13433 Sequence 13433, A
C 7	17.2	68.8	76589	7	US-10-995-561-13322 Sequence 13322, A
C 8	17.2	68.0	842	7	US-10-750-185-42851 Sequence 42851, A
C 9	17.6	70.4	1400	8	US-11-136-527-7256 Sequence 7256, Ap
C 10	17.6	70.4	1452	7	US-10-750-185-45050 Sequence 45050, A
C 11	17.6	70.4	1452	7	US-10-750-623-45050 Sequence 45050, A
C 12	17.6	70.4	1500	7	US-10-750-185-50014 Sequence 50014, A
C 13	17.6	70.4	1500	7	US-10-750-623-50014 Sequence 50014, A
C 14	17.6	70.4	1500	7	US-10-750-185-50014 Sequence 50014, A
C 15	17.6	70.4	1626	7	US-10-750-185-25035 Sequence 25035, A
C 16	17.6	70.4	1626	7	US-10-750-623-25035 Sequence 25035, A
C 17	17.6	70.4	1854	7	US-10-750-185-31957 Sequence 31957, A
C 18	17.6	70.4	1854	7	US-10-750-623-31957 Sequence 31957, A
C 19	17.6	70.4	2001	8	US-11-043-752-1189 Sequence 1189, Ap
C 20	17.6	70.4	2044	8	US-11-136-527-4085 Sequence 4085, Ap
C 21	17.6	70.4	2044	8	US-11-136-527-4085 Sequence 4085, Ap
C 22	17.6	70.4	2479	7	US-10-750-185-42601 Sequence 42601, A

Sequence 42601, A  
Sequence 54891, A  
Sequence 54891, A  
Sequence 3160, Ap  
Sequence 2922, Ap  
Sequence 62003, A  
Sequence 62003, A  
Sequence 2914, Ap  
Sequence 35994, A  
Sequence 35994, A  
Sequence 25048, A  
Sequence 25048, A  
Sequence 2030, Ap  
Sequence 13214, A  
Sequence 13214, A  
Sequence 13319, A  
Sequence 13444, A  
Sequence 1, Appli  
Sequence 82, Appli  
Sequence 36, Appli  
Sequence 13236, A  
Sequence 15, Appli  
Sequence 34, Appli  
Sequence 56, Appli  
Sequence 55, Appli  
Sequence 211, Appli  
Sequence 42575, A  
Sequence 72433, A  
Sequence 46992, A  
Sequence 46992, A  
Sequence 43537, A  
Sequence 43537, A  
Sequence 31962, A  
Sequence 31962, A  
Sequence 54596, A  
Sequence 54596, A  
Sequence 30415, A  
Sequence 33041, A  
Sequence 33041, A  
Sequence 117, App  
Sequence 13434, A  
Sequence 20, Appli  
Sequence 13293, A  
Sequence 13286, A  
Sequence 426658, A  
Sequence 40855, A  
Sequence 41669, A  
Sequence 42642, A  
Sequence 44331, A  
Sequence 50297, A  
Sequence 71235, A  
Sequence 71259, A  
Sequence 13096, A  
Sequence 13097, A  
Sequence 2807, Ap  
Sequence 31949, A  
Sequence 31949, A  
Sequence 500, App  
Sequence 58987, A  
Sequence 58987, A  
Sequence 49983, A  
Sequence 49983, A  
Sequence 33465, A  
Sequence 33465, A  
Sequence 53961, A  
Sequence 53961, A  
Sequence 130, App  
Sequence 39528, A  
Sequence 39528, A  
Sequence 38349, A  
Sequence 38349, A  
Sequence 51136, A

c 96 16 64.0 2030 7 US-10-750-623-51136 Sequence 51136, A  
c 97 16 64.0 2162 7 US-10-750-185-58286 Sequence 58286, A  
c 98 16 64.0 2162 7 US-10-750-623-58286 Sequence 58286, A  
c 99 16 64.0 2235 7 US-10-750-185-46293 Sequence 46293, A  
c 100 16 64.0 2235 7 US-10-750-623-46293 Sequence 46293, A  
c 101 16 64.0 2289 7 US-10-750-185-25692 Sequence 25692, A  
c 102 16 64.0 2289 7 US-10-750-623-25692 Sequence 25692, A  
c 103 16 64.0 2389 7 US-10-750-185-37454 Sequence 37454, A  
c 104 16 64.0 2389 7 US-10-750-623-37454 Sequence 37454, A  
c 105 16 64.0 2552 7 US-10-750-185-26948 Sequence 26948, A  
c 106 16 64.0 2552 7 US-10-750-623-26948 Sequence 26948, A  
c 107 16 64.0 2169 7 US-10-750-185-25665 Sequence 25665, A  
c 108 16 64.0 3169 7 US-10-750-623-25665 Sequence 25665, A  
c 109 16 64.0 3207 7 US-10-995-561-13317 Sequence 13317, A  
c 110 16 64.0 77246 8 US-11-124-368A-2907 Sequence 2907, Ap  
c 111 16 64.0 150450 7 US-11-112-908-54 Sequence 54, Appl  
c 112 16 64.0 165156 7 US-10-995-561-13304 Sequence 13304, A  
c 113 16 64.0 175416 8 US-11-121-086-43 Sequence 43, Appl  
c 114 16 64.0 175673 8 US-11-121-086-55 Sequence 55, Appl  
c 115 16 64.0 191343 8 US-11-112-908-53 Sequence 53, Appl  
c 116 16 64.0 212805 8 US-11-112-908-19 Sequence 19, Appl  
c 117 16 64.0 403278 7 US-10-995-561-13421 Sequence 13421, A  
c 118 16 64.0 1125000 7 US-10-995-561-13286 Sequence 13286, A  
c 119 15.8 63.2 600 7 US-10-750-185-782 Sequence 782, App  
c 120 15.8 63.2 600 7 US-10-750-623-782 Sequence 782, App  
c 121 15.8 63.2 600 8 US-11-136-527-8000 Sequence 8000, Ap  
c 122 15.8 63.2 1176 7 US-10-750-185-29230 Sequence 29230, A  
c 123 15.8 63.2 1176 7 US-10-750-623-29230 Sequence 29230, A  
c 124 15.8 63.2 1260 7 US-10-750-185-52954 Sequence 52954, A  
c 125 15.8 63.2 1260 7 US-10-750-623-52954 Sequence 52954, A  
c 126 15.8 63.2 1889 7 US-10-750-185-39394 Sequence 39394, A  
c 127 15.8 63.2 4431 7 US-10-750-623-39394 Sequence 39394, A  
c 128 15.8 63.2 4431 7 US-10-750-185-30584 Sequence 30584, A  
c 129 15.8 63.2 5415 7 US-10-750-623-30584 Sequence 30584, A  
c 130 15.8 63.2 5580 8 US-11-136-527-3904 Sequence 3904, Ap  
c 131 15.8 63.2 22 7 US-10-310-914A-426659 Sequence 426659, A  
c 132 15.6 62.4 201 7 US-10-995-561-28081 Sequence 28081, A  
c 133 15.6 62.4 201 7 US-10-995-561-76104 Sequence 76104, A  
c 134 15.6 62.4 396 7 US-10-769-744-418 Sequence 418, App  
c 135 15.6 62.4 396 8 US-11-096-191-526 Sequence 526, App  
c 136 15.6 62.4 529 8 US-11-128-061-2221 Sequence 2221, Ap  
c 137 15.6 62.4 529 8 US-11-128-061-5863 Sequence 5863, Ap  
c 138 15.6 62.4 529 8 US-11-128-049-2221 Sequence 2221, Ap  
c 139 15.6 62.4 529 8 US-11-128-049-5863 Sequence 5863, Ap  
c 140 15.6 62.4 529 8 US-10-750-185-4247 Sequence 4247, Ap  
c 141 15.6 62.4 600 7 US-10-750-623-4247 Sequence 4247, Ap  
c 142 15.6 62.4 600 7 US-10-750-185-50951 Sequence 50951, A  
c 143 15.6 62.4 732 7 US-10-750-623-50951 Sequence 50951, A  
c 144 15.6 62.4 732 7 US-11-145-703-39 Sequence 39, Appl  
c 145 15.6 62.4 985 8 US-10-750-185-55379 Sequence 55379, A  
c 146 15.6 62.4 1084 7 US-10-750-623-55379 Sequence 55379, A  
c 147 15.6 62.4 1084 7 US-10-750-185-48485 Sequence 48485, A  
c 148 15.6 62.4 1116 7 US-10-750-623-48485 Sequence 48485, A  
c 149 15.6 62.4 1116 7 US-10-750-623-48485 Sequence 48485, A  
c 150 15.6 62.4 1281 7 US-10-750-185-64025 Sequence 64025, A

## ALIGNMENTS

RESULT 1  
US-11-124-368A-2898/c  
; Sequence 2898, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09

; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2898  
; LENGTH: 100000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2898

Query Match 71.2%; Score 17.8; DB 8; Length 100000;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAACAATGACCAACTGGGC 22  
|||||  
Db 93082 AAAAACAATGACCAACTGGGC 93062

RESULT 2  
US-11-121-086-104  
; Sequence 104, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 104  
; LENGTH: 186442  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-104

Query Match 71.2%; Score 17.8; DB 8; Length 186442;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 AAACAATGACCAACTGGGCTG 24  
|||||  
Db 129637 AAAAATGACCAACTGGGCTG 129657

RESULT 3  
US-10-995-561-82214/c  
; Sequence 82214, Application US/109955561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 82214  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-82214  
Query Match 70.4%; Score 17.6; DB 7; Length 201;  
Best Local Similarity 83.3%; Pred. No. 63;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAACAATGACCAACTGGGCTGT 25  
Db 32 ACAACCAATGACCAAGTGGGCTTT 9

RESULT 4  
US-10-995-561-13495/c  
; Sequence 13495, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13495  
; LENGTH: 195235  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13495

Query Match 70.4%; Score 17.6; DB 7; Length 195235;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAACAATGACCAACTGGGCTGT 25  
Db 106040 ACAACCAATGACCAAGTGGGCTTT 106017

RESULT 5  
US-10-995-561-13421  
; Sequence 13421, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13421  
; LENGTH: 403278  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(403278)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-995-561-13421

Query Match 70.4%; Score 17.6; DB 7; Length 403278;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAACAATGACCAACTGGGCTGT 25  
Db 18688 AAACAATGATCCCAAGGCTGT 18711

RESULT 6  
US-10-995-561-13343/c  
; Sequence 13343, Application US/10995561  
; Publication No. US20050272054A1

; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13343  
; LENGTH: 83528  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13343

Query Match 69.6%; Score 17.4; DB 7; Length 83528;  
Best Local Similarity 94.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTG 19  
Db 19575 AAAAAACAATGTCCAATG 19557

RESULT 7  
US-10-995-561-13322  
; Sequence 13322, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13322  
; LENGTH: 76589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(76589)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-995-561-13322

Query Match 68.8%; Score 17.2; DB 7; Length 76589;  
Best Local Similarity 86.4%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACAATGACCAACTGGGCTGT 25  
Db 64823 AATCAATGACCAACTGTGCTGT 64844

RESULT 8  
US-10-750-185-42851  
; Sequence 42851, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185

```
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42851
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Bovine 19866881077507
US-10-750-185-42851

Query Match      68.0%; Score 17; DB 7; Length 842;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 405 AATAATCAATCACCACACTCTGCTGT 429

RESULT 9
US-10-750-623-42851
; Sequence 42851, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42851
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Bovine 19866881077507
US-10-750-623-42851

Query Match      68.0%; Score 17; DB 7; Length 842;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 405 AATAATCAATCACCACACTCTGCTGT 429

RESULT 10
US-11-136-527-7256/c
; Sequence 7256, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7256
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Bovine 19866881030841
US-10-750-185-45050

Query Match      68.0%; Score 17; DB 7; Length 1452;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 339 AAATAAARATRAYTGACTGGGCTRT 315

RESULT 11
US-11-136-527-8181/c
; Sequence 8181, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8181
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-8181

Query Match      68.0%; Score 17; DB 8; Length 1400;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
Db 339 AAATAAARATRAYTGACTGGGCTGT 315

RESULT 12
US-10-750-185-45050
; Sequence 45050, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45050
; LENGTH: 1452
; TYPE: DNA
; ORGANISM: Bovine 19866881030841
US-10-750-185-45050

Query Match      68.0%; Score 17; DB 7; Length 1452;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25
```



; Sequence 25035, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 25035  
; LENGTH: 1626  
; TYPE: DNA  
; ORGANISM: Bovine 19866808051773  
US-10-750-623-25035

Query Match 68.0%; Score 17; DB 7; Length 1626;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25  
||||| ||| ||||| |||  
Db 352 AAAAAAGAAGAAAACTGGGCAGT 328

## RESULT 18

US-10-750-185-31957  
; Sequence 31957, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 31957  
; LENGTH: 1854  
; TYPE: DNA  
; ORGANISM: Bovine 198668080691622  
US-10-750-185-31957

Query Match 68.0%; Score 17; DB 7; Length 1854;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25  
||||| ||| ||||| |||  
Db 450 AGATAACAATGACCGACTGACCTGT 474

## RESULT 19

US-10-750-623-31957  
; Sequence 31957, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth

; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 31957  
; LENGTH: 1854  
; TYPE: DNA  
; ORGANISM: Bovine 198668080691622  
US-10-750-623-31957

Query Match 68.0%; Score 17; DB 7; Length 1854;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25  
||||| ||| ||||| |||  
Db 450 AGATAACAATGACCGACTGACCTGT 474

## RESULT 20

US-11-043-752-1189  
; Sequence 1189, Application US/11043752  
; Publication No. US20060014165A1  
; GENERAL INFORMATION:  
; APPLICANT: Hakonarson, Hakon  
; APPLICANT: Gurney, Mark E.  
; APPLICANT: Halapi, Eva  
; TITLE OF INVENTION: METHODS OF DIAGNOSIS AND TREATMENT FOR  
; TITLE OF INVENTION: ASTHMA AND OTHER RESPIRATORY DISEASES BASED ON HAPLOTYPE  
; FILE REFERENCE: 2345.2044-003  
; CURRENT APPLICATION NUMBER: US/11/043,752  
; CURRENT FILING DATE: 2005-01-26  
; PRIOR APPLICATION NUMBER: PCT/US04/022446  
; PRIOR FILING DATE: 2004-07-14  
; PRIOR APPLICATION NUMBER: 60/487,072  
; PRIOR FILING DATE: 2003-07-14  
; PRIOR APPLICATION NUMBER: 60/559,611  
; PRIOR FILING DATE: 2004-04-05  
; NUMBER OF SEQ ID NOS: 4326  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1189  
; LENGTH: 2001  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-043-752-1189

Query Match 68.0%; Score 17; DB 8; Length 2001;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

Qy 1 AAAAAACAATGACCAACTGGGCTGT 25  
||||| ||| ||||| |||  
Db 336 AATAAACAAAAACCAACTGGACTAT 360

## RESULT 21

US-11-136-527-4085/c  
; Sequence 4085, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth

```

; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; PRIOR FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4085
; LENGTH: 2044
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4085

Query Match      68.0%; Score 17; DB 8; Length 2044;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
DB 983 AATAAAATGATTGACTGGGCTGT 959

RESULT 22
US-10-750-185-42601/c
; Sequence 42601, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42601
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Bovine 19866880544411
US-10-750-185-42601

Query Match      68.0%; Score 17; DB 7; Length 2479;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAAC 17
DB 1816 AAAAAACAATGACCAAC 1800

RESULT 23
US-10-750-623-42601/c
; Sequence 42601, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482

```

```

; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42601
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Bovine 19866880544411
US-10-750-623-42601

Query Match      68.0%; Score 17; DB 7; Length 2479;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAAC 17
DB 1816 AAAAAACAATGACCAAC 1800

RESULT 24
US-10-750-185-54891/c
; Sequence 54891, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54891
; LENGTH: 3109
; TYPE: DNA
; ORGANISM: Bovine 19866880656839
US-10-750-185-54891

Query Match      68.0%; Score 17; DB 7; Length 3109;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCACTGGGCTGT 25
DB 653 AAAACACAATAGACCACTGGTCTGT 629

RESULT 25
US-10-750-623-54891/c
; Sequence 54891, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482

```



```
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54891
; LENGTH: 3109
; TYPE: DNA
; ORGANISM: Bovine 19866880656839
US-10-750-623-54891

Query Match      68.0%; Score 17; DB 7; Length 3109;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| ||||| |||||
DB 653 AAAAAACAATAGACAACTGGTCTGT 629

RESULT 26
US-11-136-527-3160/c
; Sequence 3160, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3160
; LENGTH: 3590
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-3160

Query Match      68.0%; Score 17; DB 8; Length 3590;
Best Local Similarity 68.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| :|||: ||||| ||||| :|
DB 2529 AAATAAARATRAYTGACTGGGCTRT 2505

RESULT 27
US-11-124-368A-2922/c
; Sequence 2922, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2922
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2922
```

```
Query Match      68.0%; Score 17; DB 8; Length 100000;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGGCTGT 25
   ||||| ||||| ||||| ||||| |||||
DB 60543 AAAAAACATTTTCCAACTGTGCAGT 60519

RESULT 28
US-10-750-185-62003/c
; Sequence 62003, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62003
; LENGTH: 2396
; TYPE: DNA
; ORGANISM: Bovine 19866881184825
US-10-750-185-62003

Query Match      67.2%; Score 16.8; DB 7; Length 2396;
Best Local Similarity 90.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACAATGACCAACTGGGCT 23
   ||||| ||||| ||||| |||||
DB 147 AAAAAATGACCAGCTGGGCT 128

RESULT 29
US-10-750-623-62003/c
; Sequence 62003, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62003
; LENGTH: 2396
; TYPE: DNA
; ORGANISM: Bovine 19866881184825
US-10-750-623-62003

Query Match      67.2%; Score 16.8; DB 7; Length 2396;
Best Local Similarity 90.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 4 AAACAATGACCACTGGGCT 23  
DB 147 AAAAATGACCACTGGGCT 128

RESULT 30  
US-11-124-368A-2914  
; Sequence 2914, Application US/111124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; PRIOR FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2914  
; LENGTH: 73404  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2914

Query Match 67.2%; Score 16.8; DB 8; Length 73404;  
Best Local Similarity 90.8%; Pred. No. 3.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACAATGACCACTGGGC 22  
DB 22412 AAAAATGACCACTGGGC 22431

RESULT 31  
US-10-750-185-35994/c  
; Sequence 35994, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35994  
; LENGTH: 1601  
; TYPE: DNA  
; ORGANISM: Bovine 19866881606127  
US-10-750-185-35994

Query Match 66.4%; Score 16.6; DB 7; Length 1601;  
Best Local Similarity 82.6%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCACTGGGCTG 24  
DB 1402 AAAACATCAGTCACCTGGGCTG 1380

RESULT 32  
US-10-750-623-35994/c  
; Sequence 35994, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35994  
; LENGTH: 1601  
; TYPE: DNA  
; ORGANISM: Bovine 19866881606127  
US-10-750-623-35994

Query Match 66.4%; Score 16.6; DB 7; Length 1601;  
Best Local Similarity 82.6%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACAATGACCACTGGGCTG 24  
DB 1402 AAAACATCAGTCACCTGGGCTG 1380

RESULT 33  
US-10-750-185-25048  
; Sequence 25048, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25048  
; LENGTH: 3558  
; TYPE: DNA  
; ORGANISM: Bovine 19866880528126  
US-10-750-185-25048

Query Match 66.4%; Score 16.6; DB 7; Length 3558;  
Best Local Similarity 82.6%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAACAATGACCACTGGGCT 23  
DB 2007 AAAAATGACCACTGGGCT 2029

RESULT 34

```

US-10-750-623-25048
; FILE REFERENCE: CL001559
; Sequence 25048, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25048
; LENGTH: 3558
; TYPE: DNA
; ORGANISM: Bovine 19866880528126
US-10-750-623-25048

Query Match      66.4%; Score 16.6; DB 7; Length 3558;
Best Local Similarity 82.6%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCT 23
   ||||| ||||| ||||| |||||
Db 2007 AAAAAACACTGACAACTGTGT 2029

RESULT 35
US-11-136-527-2030/c
; Sequence 2030, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2030
; LENGTH: 4095
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2030

Query Match      66.4%; Score 16.6; DB 8; Length 4095;
Best Local Similarity 82.6%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCT 23
   ||||| ||||| ||||| |||||
Db 3110 AAAACACACGTCAACTGGGCT 3088

RESULT 36
US-10-995-561-13214/c
; Sequence 13214, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13214
; LENGTH: 19277
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13214

Query Match      66.4%; Score 16.6; DB 7; Length 19277;
Best Local Similarity 82.6%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAATGACCAACTGGGCTG 24
   ||||| ||||| ||||| |||||
Db 13487 AAAAAACATGACTCACTGGCTG 13465

RESULT 37
US-10-995-561-13319/c
; Sequence 13319, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13319
; LENGTH: 27509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13319

Query Match      66.4%; Score 16.6; DB 7; Length 27509;
Best Local Similarity 82.6%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAATGACCAACTGGGCT 23
   ||||| ||||| ||||| |||||
Db 15454 AAAAAAACCAGCTGGGCT 15432

RESULT 38
US-10-995-561-13444
; Sequence 13444, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13444
; LENGTH: 29618
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13444

Query Match      66.4%; Score 16.6; DB 7; Length 29618;
Best Local Similarity 82.6%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 AAAAAAATGACCAACTGGGCTG 24  
Db 25604 AAAAAAATGACCAACTGGGCTG 25626

## RESULT 39

US-10-857-780-1/c  
; Sequence 1, Application US/10857780  
; Publication No. US20050272043A1  
; GENERAL INFORMATION:  
; APPLICANT: ROTH, RICHARD B.  
; APPLICANT: BRAUN, ANDREAS  
; APPLICANT: KAMMERER, STEFAN M.  
; APPLICANT: NELSON, MATTHEW ROBERTS  
; APPLICANT: RENELAND, RIKARD HENRY  
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS  
; FILE OF INVENTION: THEREOF  
; FILE REFERENCE: SEQ-4069-CP  
; CURRENT APPLICATION NUMBER: US/10/857,780  
; CURRENT FILING DATE: 2004-05-28  
; PRIOR APPLICATION NUMBER: 10/723,681  
; PRIOR FILING DATE: 2003-11-25  
; PRIOR APPLICATION NUMBER: 60/490,234  
; PRIOR FILING DATE: 2003-07-24  
; PRIOR APPLICATION NUMBER: 60/525,239  
; PRIOR FILING DATE: 2003-11-25  
; NUMBER OF SEQ ID NOS: 4962  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 92600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (22393)..(22394)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (24675)..(24676)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (31942)..(31942)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (32965)..(32966)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33724)..(33724)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-857-780-1

Query Match 66.4%; Score 16.6; DB 7; Length 92600;  
Best Local Similarity 82.6%; Pred. No. 4.6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCT 23  
Db 32635 AAAAAAATGACCAACTGGGCT 32613

## RESULT 40

US-10-775-169-82  
; Sequence 82, Application US/10775169  
; Publication No. US20050287532A9  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dörner, Andrew

; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 82  
; LENGTH: 127917  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-82

Query Match 66.4%; Score 16.6; DB 7; Length 127917;  
Best Local Similarity 82.6%; Pred. No. 4.8e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCT 23  
Db 90851 AAAAAAATGACCAACTGGGCT 90873

## RESULT 41

US-11-121-086-36  
; Sequence 36, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 36  
; LENGTH: 172649  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-36

Query Match 66.4%; Score 16.6; DB 8; Length 172649;  
Best Local Similarity 82.6%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACTGGGCT 23  
Db 12905 AAAAAAATGACCAACTGGGCT 12927

## RESULT 42

US-10-995-561-13236  
; Sequence 13236, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13236  
; LENGTH: 305312  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(305312)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-4)  
US-10-995-561-12336

Query Match 66.4%; Score 16.6; DB 7; Length 305312;  
Best Local Similarity 82.6%; Pred. No. 5.2e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAATGACCAACTGGCTG 24  
||||| ||||| ||||| ||||| |||||  
Db 216461 AAAAAATGACCACTGCTG 216483

RESULT 43  
US-10-999-208-15/C  
; Sequence 15, Application US/10999208  
; Publication No. US20060008448A1  
; GENERAL INFORMATION:  
; APPLICANT: XU, MINZHEN  
; APPLICANT: HUMPHREYS, ROBERT  
; TITLE OF INVENTION: INHIBITION OF LI EXPRESSION IN MAMMALIAN CELLS  
; FILE REFERENCE: REH-2016US01  
; CURRENT APPLICATION NUMBER: US/10/999,208  
; PRIOR FILING DATE: 2004-11-29  
; PRIOR APPLICATION NUMBER: 10/127,347  
; PRIOR FILING DATE: 2002-04-22  
; PRIOR APPLICATION NUMBER: 10/054,387  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: 09/205,995  
; PRIOR FILING DATE: 1998-12-04  
; PRIOR APPLICATION NUMBER: 09/036,746  
; PRIOR FILING DATE: 1998-03-09  
; PRIOR APPLICATION NUMBER: 08/661,627  
; PRIOR FILING DATE: 1996-06-11  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 15  
; LENGTH: 57  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-10-999-208-15

Query Match 65.6%; Score 16.4; DB 6; Length 57;  
Best Local Similarity 94.4%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAATGACCACTG 19  
||||| ||||| ||||| ||||| |||||  
Db 57 AAAAAATGACCACTG 40

RESULT 44  
US-11-121-438-34  
; Sequence 34, Application US/11121438  
; Publication No. US20060014173A1  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: PR/SET- Domain Containing Nucleic Acids, Antibodies and Methods of Use  
; FILE REFERENCE: P-LJ 5301  
; CURRENT APPLICATION NUMBER: US/11/121,438  
; PRIOR FILING DATE: 2005-05-03  
; PRIOR APPLICATION NUMBER: US/10/200,012  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: US 09/910,478  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 1490  
; TYPE: DNA

; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (64)...(1473)  
US-11-121-438-34

Query Match 65.6%; Score 16.4; DB 8; Length 1490;  
Best Local Similarity 94.4%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AACCAATGACCAACTGGGC 22  
||||| ||||| ||||| ||||| |||||  
Db 1165 AACCAATGACCAACTGGGC 1182

RESULT 45  
US-11-112-908-56/c  
; Sequence 56, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 56  
; LENGTH: 150468  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-56

Query Match 65.6%; Score 16.4; DB 8; Length 150468;  
Best Local Similarity 94.4%; Pred. No. 6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAATGACCAACT 18  
||||| ||||| ||||| ||||| |||||  
Db 3315 AAAAAAATGACCAACT 3298

RESULT 46  
US-11-112-908-55/c  
; Sequence 55, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 55

```
; LENGTH: 193789
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-55

Query Match      65.6%; Score 16.4; DB 8; Length 193789;
Best Local Similarity 94.4%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACT 18
Db 53500 AAAAAACAATGACCAACT 53483

RESULT 47
US-11-117-187-211
; Sequence 211, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-211

Query Match      65.6%; Score 16.4; DB 8; Length 1082144;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACT 18
Db 824756 AAAAAACAATGACCAACT 824773

RESULT 48
US-10-995-561-42575
; Sequence 42575, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42575
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-42575

Query Match      64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGC 22
Db 170 AAAAAACAATGACCAACTGGGC 190

us-10-719-900-5.rnpbn

RESULT 49
US-10-995-561-72433
; Sequence 72433, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72433
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-72433

Query Match      64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACAATGACCAACTGGGC 22
Db 170 AAAAAACAATGACCAACTGGGC 190

RESULT 50
US-10-750-185-46992
; Sequence 46992, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46992
; LENGTH: 1030
; TYPE: DNA
; ORGANISM: Bovine 19866880867784
US-10-750-185-46992

Query Match      64.8%; Score 16.2; DB 7; Length 1030;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACAATGACCAACTGGG 21
Db 813 AAAAAACAATGACCAACTGGG 833

Search completed: February 3, 2006, 16:19:52
Job time : 348.111 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttggaaggagactca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_ov.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_ets.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21.4	85.6	110000	14	AC152182_1	Continuation (2 of
2	21.4	85.6	110000	14	AC155165_2	Continuation (3 of
3	21.4	85.6	200774	9	AC116324	AC116324 Mus muscu
4	20.8	83.2	844	9	BC048869	BC048869 Mus muscu
5	20.8	83.2	859	9	BC024624	BC024624 Mus muscu
6	20.8	83.2	870	9	MUSEGFBPM	M17962 Mouse major
7	20.8	83.2	878	9	BC026378	BC026378 Mus muscu
8	20.2	80.8	412	5	AY493302	AY493302 Anas plat
9	20.2	80.8	158857	14	AC165233	AC165233 Mus muscu
10	20.2	80.8	168803	8	AC142350	AC142350 Pan trogl
11	20.2	80.8	189158	8	AC147666	AC147666 Mus muscu
12	20.2	80.8	196620	14	AC145502	AC145502 Canis fam
13	20.2	80.8	210317	9	AC102363	AC102363 Mus muscu
14	20.2	80.8	214385	14	AC153790	AC153790 Mus muscu
15	20.2	80.8	215892	14	AC152260	AC152260 Bos tauru
16	20.2	80.8	230954	14	AC116092	AC116092 Rattus no
17	19.8	79.2	121010	8	AC140677	AC140677 Homo sapi
18	19.8	79.2	163805	8	AC114959	AC114959 Homo sapi

AC156954	Mus muscu	9	AC156954	79.2	179140	19.8	19.8	C 19
AC131565	Homo sapi	8	AC131565	79.2	181283	19.8	19.8	C 20
AF129075	Homo sapi	8	AF129075	79.2	190837	19.8	19.8	C 21
BS000157	Pan trogl	5	BS000157	79.2	192836	19.8	19.8	C 22
AL954132	Zebrafish	8	AL954132	79.2	197745	19.8	19.8	C 23
BS000171	Pan trogl	8	BS000171	79.2	218680	19.8	19.8	C 24
AC153628	Mus muscu	8	AC153628	79.2	219322	19.8	19.8	C 25
AL161324	Homo sapi	8	HS21C049	79.2	340000	19.8	19.8	C 26
AC018002	Drosophil	14	AC018002	77.6	73237	19.4	19.4	C 27
AC005442	Drosophil	14	AC005442	77.6	90788	19.4	19.4	C 28
AC009250	Drosophil	2	AC009250	77.6	170362	19.4	19.4	C 29
AC090872	Drosophil	2	AC090872	77.6	170362	19.4	19.4	C 30
AC152332	Bos tauru	14	AC152332	77.6	178461	19.4	19.4	C 31
AE003653	Drosophil	2	AE003653	77.6	247634	19.4	19.4	C 32
X00472	Mouse mRNa	9	MMTAM1	76.8	428	19.2	19.2	C 33
BV161004	RPAMMSEQO	10	BV161004	76.8	627	19.2	19.2	C 34
X03560	Rat tissue	9	RNTKALRC	76.8	689	19.2	19.2	C 35
S48142	T-kininogen	9	S48142	76.8	777	19.2	19.2	C 36
X01389	Mouse mRNa	9	MMNGFB	76.8	849	19.2	19.2	C 37
JO0758	rat pancrea	9	RATEPKJK	76.8	867	19.2	19.2	C 38
M11563	Rat submaxi	9	RATKALPS	76.8	873	19.2	19.2	C 39
BC078784	Rattus no	9	BC078784	76.8	879	19.2	19.2	C 40
Z37432	p.hysosipifo	15	PHBBL52	76.8	1420	19.2	19.2	C 41
AB033705	Homo sapi	8	AB036705	76.8	46235	19.2	19.2	C 42
AC101212	Mus muscu	14	AC101212	76.8	59388	19.2	19.2	C 43
Continuation (3 of		14	AC123415_2	76.8	110000	19.2	19.2	C 44
AC159696	Bos tauru	14	AC159696	76.8	125379	19.2	19.2	C 45
AP000405	Homo sapi	14	AP000405	76.8	141675	19.2	19.2	C 46
AC073631	Homo sapi	14	AC073631	76.8	155877	19.2	19.2	C 47
AC161483	Mus muscu	14	AC161483	76.8	161497	19.2	19.2	C 48
AC114655	Mus muscu	9	AC114655	76.8	162542	19.2	19.2	C 49
AL590406	Human DNA	8	AL590406	76.8	164831	19.2	19.2	C 50
AL590409	Homo sapi	14	AL590409	76.8	167305	19.2	19.2	C 51
AP000487	Homo sapi	8	AP000487	76.8	177076	19.2	19.2	C 52
AC012152	Homo sapi	8	AC012152	76.8	180194	19.2	19.2	C 53
EX927064	Zebrafish	5	EX927064	76.8	181758	19.2	19.2	C 54
AC129433	Rattus no	14	AC129433	76.8	186135	19.2	19.2	C 55
AC125349	Mus muscu	9	AC125349	76.8	194060	19.2	19.2	C 56
CT010451	Mus muscu	14	CT010451	76.8	194513	19.2	19.2	C 57
AC160017	Mus muscu	14	AC160017	76.8	219235	19.2	19.2	C 58
AC111047	Mus muscu	9	AC111047	76.8	230041	19.2	19.2	C 59
AC094376	Rattus no	14	AC094376	76.8	238301	19.2	19.2	C 60
AC099189	Rattus no	14	AC099189	76.8	238635	19.2	19.2	C 61
AC106481	Rattus no	14	AC106481	76.8	239343	19.2	19.2	C 62
AC125913	Rattus no	14	AC125913	76.8	240633	19.2	19.2	C 63
AC135007	Rattus no	14	AC135007	76.8	248240	19.2	19.2	C 64
AC163166	Bos tauru	14	AC163166	76.8	249984	19.2	19.2	C 65
AC098014	Rattus no	14	AC098014	76.8	264090	19.2	19.2	C 66
AC159080	Bos tauru	14	AC159080	76.8	299322	19.2	19.2	C 67
AC159080	Bos tauru	14	AC159080	76.8	299322	19.2	19.2	C 68
AC103473	Rattus no	14	AC103473	76.0	230787	19	19	C 69
AC130579	Rattus no	14	AC130579	76.0	284657	19	19	C 70
G01621	chicken STS	10	G01621	75.2	376	18.8	18.8	C 71
BC090788	Danio rer	5	BC090788	75.2	1255	18.8	18.8	C 72
AF538351	Lolium pe	15	AF538351	75.2	4415	18.8	18.8	C 73
AC101324	Mus muscu	14	AC101324	75.2	67665	18.8	18.8	C 74
AC018777	Homo sapi	14	AC018777	75.2	70605	18.8	18.8	C 75
AC021273	Homo sapi	14	AC021273	75.2	75594	18.8	18.8	C 76
AC120305	Homo sapi	8	AC120305	75.2	95234	18.8	18.8	C 77
Continuation (3 of		14	AC112124_2	75.2	110000	18.8	18.8	C 78
Continuation (10 o		15	AE017345_09	75.2	110000	18.8	18.8	C 79
AC122419	Mus muscu	9	AC122419	75.2	131907	18.8	18.8	C 80
AL671905	Mouse DNA	9	AL671905	75.2	140012	18.8	18.8	C 81
AC025324	Homo sapi	14	AC025324	75.2	148871	18.8	18.8	C 82
AL627406	Mouse DNA	9	AL627406	75.2	155214	18.8	18.8	C 83
AC025507	Homo sapi	14	AC025507	75.2	156207	18.8	18.8	C 84
AC124476	Mus muscu	9	AC124476	75.2	169031	18.8	18.8	C 85
BX088597	Zebrafish	5	BX088597	75.2	172098	18.8	18.8	C 86
AC094105	Homo sapi	8	AC094105	75.2	186314	18.8	18.8	C 87
AL929443	Mouse DNA	9	AL929443	75.2	189821	18.8	18.8	C 88
AL929443	Mouse DNA	9	AL929443	75.2	195117	18.8	18.8	C 89
AC136065	Rattus no	14	AC136065	75.2	197694	18.8	18.8	C 90
AC023794	Homo sapi	8	AC023794	75.2	198488	18.8	18.8	C 91

c 92	18.8	75.2	215728	14	AC099112	AC099112 Rattus no
c 93	18.8	75.2	225152	9	AC132454	AC132454 Mus muscu
c 94	18.8	75.2	237165	9	AC112948	AC112948 Mus muscu
c 95	18.8	75.2	243674	14	AC098471	AC098471 Rattus no
c 96	18.8	75.2	249781	14	AC098954	AC098954 Rattus no
c 97	18.8	75.2	294415	14	AC110321	AC110321 Rattus no
c 98	18.6	74.4	520	10	BV279664	BV279664 S232P644R
c 99	18.6	74.4	558	10	BV401108	BV401108 S229P6164
c 100	18.6	74.4	770	10	BV598075	BV598075 S217P6671
c 101	18.6	74.4	1139	9	MUSLRT221	M12819 Mouse f-cal
c 102	18.6	74.4	1268	6	A95291	A95291 Sequence 22
c 103	18.6	74.4	2424	5	BC075293	BC075293 Xenopus t
c 104	18.6	74.4	5310	9	MLVYT22	Y00157 Murine Lys-
c 105	18.6	74.4	6316	6	MUSNHLTY2	M22064 Mouse MHC c
c 106	18.6	74.4	9904	6	CQ728820	CQ728820 Sequence
c 107	18.6	74.4	10327	8	AB002309	AB002309 Human mRN
c 108	18.6	74.4	10343	8	UI17195	UI17195 Homo sapien
c 109	18.6	74.4	45154	14	AC164871	AC164871 Sorex ara
c 110	18.6	74.4	61004	8	AL357566	AL357566 Human DNA
c 111	18.6	74.4	64849	5	AC098643	AC098643 Takifugu
c 112	18.6	74.4	68166	14	AC084712	AC084712 Homo sapi
c 113	18.6	74.4	71266	14	AP007726	AP007726 Lotus cor
c 114	18.6	74.4	71455	14	AC146319	AC146319 Takifugu
c 115	18.6	74.4	79585	15	AP004908	AP004908 Lotus cor
c 116	18.6	74.4	82859	14	AP007748	AP007748 Lotus cor
c 117	18.6	74.4	83607	14	AC101668	AC101668 Mus muscu
c 118	18.6	74.4	86050	8	AL359647	AL359647 Human DNA
c 119	18.6	74.4	92863	11	GGBL0C05	AL023516 Gallus ga
c 120	18.6	74.4	103019	5	AC091292	AC091292 Takifugu
c 121	18.6	74.4	105245	14	AP007325	AP007325 Lotus cor
c 122	18.6	74.4	107745	8	AL139042	AL139042 Human DNA
c 123	18.6	74.4	108535	14	AC150835	AC150835 Takifugu
c 124	18.6	74.4	110000	14	AC092450	AC092450 Homo sapi
c 125	18.6	74.4	110000	14	AC095357_1	Continuation (2 of
c 126	18.6	74.4	110000	14	CT010578_1	Continuation (2 of
c 127	18.6	74.4	113027	8	AC090669	AC090669 Homo sapi
c 128	18.6	74.4	115040	15	AC124954	AC124954 Medicago
c 129	18.6	74.4	120302	15	AC152403	AC152403 Medicago
c 130	18.6	74.4	121466	8	AC105310	AC105310 Homo sapi
c 131	18.6	74.4	121479	14	AC160837	AC160837 Medicago
c 132	18.6	74.4	125483	14	AC144888	AC144888 Bos tauru
c 133	18.6	74.4	127360	14	AC099409	AC099409 Felis cat
c 134	18.6	74.4	131264	14	AC104198	AC104198 Mus muscu
c 135	18.6	74.4	134569	8	AC105123	AC105123 Homo sapi
c 136	18.6	74.4	135855	14	AC098701	AC098701 Felis cat
c 137	18.6	74.4	137554	14	AC127618	AC127618 Rattus no
c 138	18.6	74.4	139062	15	AC133863	AC133863 Medicago
c 139	18.6	74.4	144439	15	AC147008	AC147008 Medicago
c 140	18.6	74.4	147908	5	AC091293	AC091293 Takifugu
c 141	18.6	74.4	150539	14	CR336398	CR336398 Danio rer
c 142	18.6	74.4	157145	5	BX293547	BX293547 Zebrafish
c 143	18.6	74.4	157842	8	AC092825	AC092825 Homo sapi
c 144	18.6	74.4	157981	14	AC144693	AC144693 Sus scrof
c 145	18.6	74.4	159314	14	AC134321	AC134321 Felis cat
c 146	18.6	74.4	159793	14	AC120687	AC120687 Rattus no
c 147	18.6	74.4	160359	8	AC069272	AC069272 Homo sapi
c 148	18.6	74.4	161781	14	CR376800	CR376800 Danio rer
c 149	18.6	74.4	162907	14	AC026074	AC026074 Homo sapi
c 150	18.6	74.4	164612	9	AC124697	AC124697 Mus muscu

## ALIGNMENTS

RESULT 1	AC152182_1	Sequence split into 5 fragments	LOCUS AC152182	Accession	AC152182
WPCOMMENT	Fragment Name	Begin	End		
	AC152182_0	1	110000		
	AC152182_1	100001	210000		
	AC152182_2	200001	310000		
	AC152182_3	300001	410000		
Continuation (2 of 5) of AC152182 from base 100001 (AC152182 Mus musculus chromosome 16					
400001 452855					
Query Match 85.6%; Score 21.4; DB 14; Length 110000;					
Best Local Similarity 95.7%; Pred. No. 16; Length 110000;					
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY 2 AAAAAACACTTGGAGGGAGACTC 24					
Db 29883 AAAAACTCTTGGAGGGAGACTC 29905					
RESULT 2					
AC155165_2/c					
WPCOMMENT					
Sequence split into 6 fragments LOCUS AC155165 Accession AC155165					
Fragment Name Begin End					
AC155165_0 1 110000					
AC155165_1 100001 210000					
AC155165_2 200001 310000					
AC155165_3 300001 410000					
AC155165_4 400001 510000					
AC155165_5 500001 551847					
Continuation (3 of 6) of AC155165 from base 200001 (AC155165 Mus musculus chromosome 16					
Query Match 85.6%; Score 21.4; DB 14; Length 110000;					
Best Local Similarity 95.7%; Pred. No. 16;					
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY 2 AAAAAACACTTGGAGGGAGACTC 24					
Db 59362 AAAAACTCTTGGAGGGAGACTC 59340					
RESULT 3					
AC116324/c					
LOCUS					
AC116324 Mus musculus BAC clone RP23-282P1 from 16, complete sequence.					
DEFINITION					
AC116324					
ACCESSION					
AC116324.6 GI:29244826					
VERSION					
HTG.					
KEYWORDS					
SOURCE					
Mus musculus (house mouse)					
ORGANISM					
Mus musculus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;					
Sciurognathi; Muridea; Muridae; Murinae; Mus.					
REFERENCE					
AUTHORS					
Goyea,E., Haglund,K. and Dignan,G.					
TITLE					
The sequence of Mus musculus BAC clone RP23-282P1					
JOURNAL					
Unpublished (2001)					
REFERENCE					
AUTHORS					
Wilson,R.					
TITLE					
Sequencing of Mus musculus					
JOURNAL					
Unpublished (2001)					
REFERENCE					
AUTHORS					
McPherson,J.D. and Waterston,R.H.					
TITLE					
Direct Submission					
JOURNAL					
Submitted (26-MAR-2003) Genome Sequencing Center, 4444 Forest Park					
REFERENCE					
AUTHORS					
Wilson,R.					
TITLE					
Direct Submission					
JOURNAL					
Submitted (08-NOV-2003) Department of Genetics, Washington					

## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 26, 2003 this sequence version replaced gi:24943072.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: M\_BA0282P01  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

## FEATURES

source	Location/Qualifiers
	1. .200774
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="16"
	/map="16"
	/clone="RP23-282P1"
	/clone.lib="RPCI-23"
repeat_region	52..251
	/rpt_family="B2"
repeat_region	754..1225
	/rpt_family="L1"
repeat_region	1230..1461
	/rpt_family="L1"
repeat_region	1835..2109
	/rpt_family="B4"
repeat_region	2159..2276
	/rpt_family="MaLR"
repeat_region	2301..2473
	/rpt_family="B2"
repeat_region	2665..2741
	/rpt_family="Alu"
repeat_region	2742..2878
	/rpt_family="Alu"
repeat_region	2879..2966
	/rpt_family="Alu"
repeat_region	4183..4325
	/rpt_family="ERVK"
repeat_region	4455..4567
	/rpt_family="ERVK"
repeat_region	4732..4930
	/rpt_family="MaLR"
repeat_region	4954..5144
	/rpt_family="L1"
repeat_region	5132..6177
	/rpt_family="L1"
repeat_region	6178..6398
	/rpt_family="MaLR"
repeat_region	6732..6924
	/rpt_family="B4"
repeat_region	6930..7001
	/rpt_family="B4"
repeat_region	7019..7097
	/rpt_family="ID"
repeat_region	7169..7632
	/rpt_family="L1"
repeat_region	7792..8195
	/rpt_family="MaLR"
repeat_region	8458..8725
	/rpt_family="B4"
repeat_region	8944..8998
	/rpt_family="B2"
repeat_region	8999..9451
	/rpt_family="ERV1"
repeat_region	9452..9592
	/rpt_family="B2"
repeat_region	9932..10075
	/rpt_family="B4"
repeat_region	10765..10875
	/rpt_family="B4"
repeat_region	11883..12045
	/rpt_family="B4"
repeat_region	12128..12498
	/rpt_family="RMER17C"
repeat_region	12546..12596
	/rpt_family="B4"
repeat_region	16707..16777
	/rpt_family="Alu"
repeat_region	17062..17381
	/rpt_family="B4"
repeat_region	19405..19807
	/rpt_family="L1"
repeat_region	19806..20003
	/rpt_family="L1"
repeat_region	21347..21727
	/rpt_family="MaLR"
repeat_region	21887..22098
	/rpt_family="MaLR"
repeat_region	22803..22853
	/rpt_family="B4"
repeat_region	22971..23430
	/rpt_family="ERVK"
repeat_region	23426..23593
	/rpt_family="B4"
repeat_region	23816..23985
	/rpt_family="B4"
repeat_region	23992..24131
	/rpt_family="ERV1"
repeat_region	25013..25148
	/rpt_family="Alu"
repeat_region	25661..25816
	/rpt_family="B4"
repeat_region	26032..26117
	/rpt_family="B4"
repeat_region	26797..26938
	/rpt_family="Alu"
repeat_region	27033..27194
	/rpt_family="L1"
repeat_region	27215..27721
	/rpt_family="L1"
repeat_region	28012..28220
	/rpt_family="ERV1"
repeat_region	28221..28630
	/rpt_family="MaLR"
repeat_region	28631..28777

```

/rpt_family="ERV1"
repeat_region 29116..29394
/rpt_family="L1"
repeat_region 29651..29750
/rpt_family="Alu"
repeat_region 29769..29865
/rpt_family="L1"
repeat_region 30023..30248
/rpt_family="L1"
repeat_region 30249..30661
/rpt_family="ERVK"
repeat_region 30662..30911
/rpt_family="L1"
repeat_region 31018..31183
/rpt_family="B4"
repeat_region 33853..33941
/rpt_family="B4"

Query Match 85.6%; Score 21.4; DB 9; Length 200774;
Best Local Similarity 95.7%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGGAGACTC 24
|||||
Db 84737 AAAAACTTGGAGGGAGACTC 84715

RESULT 4
BC048869
LOCUS
DEFINITION Mus musculus kallikrein 9, mRNA (CDNA clone MGC:59025
IMAGE:4164261), complete cds.
ACCESSION BC048869.1 GI:29179502
VERSION BC048869.1
KEYWORDS MGC.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 844)
Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
Collins FS, Wagner CM, Shenmen CW, Schuler GD, Altschul SF, Zeeberg
B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
Max SI, Wang J, Haieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
Schetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
Helton E, Kettman M, Madan A, Rodrigues S, Sanchez A, Whiting M,
Madan A, Jung AC, Shevchenko Y, Bouffard GG, Blakesley RW,
Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
Smalilus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 844)
DIRECTOR MGC Project.
Direct Submission
Submitted (14-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nigr.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 108 Row: C Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21426850.

FEATURES
Location/Qualifiers
1..844
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:59025 IMAGE:4164261"
/tissue_type="Salivary gland, 10 week old female mouse"
/clone_lib="NCI CGAP_SG2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..844
/gene="Klk9"
/note="synonyms: Klklb9, mGk-9"
/db_xref="GeneID:13648"
/db_xref="MGI:95293"
1..786
/gene="Klk9"
/codon_start=1
/product="kallikrein 9"
/protein_id="AAH4869.1"
/db_xref="GI:29179503"
/db_xref="GeneID:13648"
/db_xref="MGI:95293"
/translat="NRFLFLSLGLGIDAAPVHRSIVTGGFKCEKNSOPWHVAVYR
YNEYICGVLIDANWLTAAHCYVEENKSLGKNLYEPPSAQHLVSKSLFHPGYN
RSLRHRHREYDYSNLMRLRLSKPADIDIVXPALPTEPKLSTCLASGSGST
TPFFQNAKDLCQVNLKLLPNEDCGKAHIEKVTVDVNLCAGETDGGDKCKGDSGGPLI
CDGVLOQITSGFTPCGPKPKGVYTKLIKFTSWIKDTMAKNL"

ORIGIN
Query Match 83.2%; Score 20.8; DB 9; Length 844;
Best Local Similarity 91.7%; Pred. No. 62;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGGAGACTCA 25
|||||
Db 616 AAGACACTTGCAGGGAGACTCA 639
|||||

RESULT 5
BC024624
LOCUS
DEFINITION Mus musculus kallikrein 9, mRNA (CDNA clone MGC:25400
IMAGE:4914236), complete cds.
ACCESSION BC024624
VERSION BC024624.1 GI:19353194
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

```

```
1 (bases 1 to 859)
REFERENCE
AUTHORS
Strausberg,R.D., Colling,F.S., Wagner,L.H., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Colling,F.S., Wagner,L.H., Shemen,C.M., Schuler,G.D.,
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 859)
REFERENCE
AUTHORS
Strausberg,R.
Direct Submission
Submitted (01-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guparatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 31 Row: p Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21426850.
FEATURES
source
Location/Qualifiers
1..859
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:25400 IMAGE:4914236"
/tissue_type="Salivary gland, 10 week old female mouse"
/clone_lib="NCI CGAP SG2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
gene
1..859
/gene="Klk9"
/note="synonym: mgk-9"
/db_xref="GeneID:13648"
/db_xref="MGI:95293"
16..801
/gene="Klk9"
/codon_start=1
/product="kallikrein 9"
/protein_id="AAH24624.1"
/db_xref="GI:19353195"
/db_xref="GeneID:13648"
CDS
BC026378
Mus musculus kallikrein 1, mRNA (cdna clone MGC:31128
IMAGE:4165617), complete cds.
BC026378
BC026378.1 GI:20071344
```

```
/db_xref="MGI:95293"
/translation="MRFLILFLALSLGGIDAAPVHRSRVGGFKCEKNSQPMHVAVYR
YNEYICGGVLLDANVLTAAHCYVEENKVLGKNLYEEPSAQHRLVSKSFHPGYN
RSLHRNHHPEDYSDNLMRLRLSKPADITDVVKPIALPTEEPKLGSTCLASGWSGT
TPFKFONAKDLQCVNLKLLPNEDCGKAHEKVTYVLMCAGETDGGDKTCKGDSGGPLI
CDGVLQGITSWGFTPGCGPKPGVYTKLILKFTSWIKDTMAKNL"
ORIGIN
Query Match 83.2%; Score 20.8; DB 9; Length 859;
Best Local Similarity 91.7%; Pred. No. 62;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAACACTTGGAAAGGAGACTCA 25
DB 631 AAAGACACTTGCAGGGAGACTCA 654
RESULT 6
MUSEGFBPM 870 bp mRNA linear ROD 12-JUN-1993
LOCUS Mouse major epidermal growth factor binding protein (EGF-BP) mRNA,
DEFINITION complete cds.
ACCESSION M17962
VERSION M17962.1 GI:192997
KEYWORDS epidermal growth factor binding protein.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 870)
AUTHORS Blaber,M., Isackson,P.J. and Bradshaw,R.A.
TITLE A complete cDNA sequence for the major epidermal growth factor
binding protein in the male mouse submandibular gland
JOURNAL Biochemistry 26 (21), 6742-6749 (1987)
PUBMED 332386
COMMENT Original source text: Mouse (Swiss Webster strain) adult male
submandibular gland, cDNA to mRNA, clone MB2-20A.
FEATURES
source
Location/Qualifiers
1..870
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
<1..870
/product="EGF-BP mRNA"
CDS
37..822
/note="major epidermal growth factor binding protein"
/codon_start=1
/protein_id="AAA37541.1"
/db_xref="GI:309212"
/translation="NRFLILFLALSLGGIDAAPVHRSRVGGFKCEKNSQPMHVAVYR
YNEYICGGVLLDANVLTAAHCYVEENKVLGKNLYEEPSAQHRLVSKSFHPGYN
RSLHRNHHPEDYSDNLMRLRLSKPADITDVVKPIALPTEEPKLGSTCLASGWSGT
TPFKFONAKDLQCVNLKLLPNEDCGKAHEKVTYVLMCAGETDGGDKTCKGDSGGPLI
CDGVLQGITSWGFTPGCGPKPGVYTKLILKFTSWIKDTMAKNL"
ORIGIN
Query Match 83.2%; Score 20.8; DB 9; Length 870;
Best Local Similarity 91.7%; Pred. No. 62;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAACACTTGGAAAGGAGACTCA 25
DB 652 AAAGACACTTGCAGGGAGACTCA 675
RESULT 7
BC026378
Mus musculus kallikrein 1, mRNA (cdna clone MGC:31128
IMAGE:4165617), complete cds.
BC026378
BC026378.1 GI:20071344
```

KEYWORDS	MGC.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 878) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huylk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 878)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a> Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a> Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 43 Row: d Column: 13 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8659567. Location/Qualifiers 1..878 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:31128 IMAGE:4165617" /tissue_type="Salivary gland, 10 week old female mouse" /clone_lib="NCI CGAP_SG2" /lab_host="DH10B" /notes="Vector: pCMV-SPORT6"
source	1..878 /gene="Klk1" /notes="Synonyms: Klk6, mGK-1, TK, mK1" /db_xref="GeneID:16623" /db_xref="MGI:892019" 20..805 /gene="Klk1" /codon_start=1
gene	AC165233 Mus musculus chromosome 8 clone RP23-30702 map 8, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces. AC165233 AC165233.4 GI:71143318 HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN. Mus musculus (house mouse)
CDS	

Query Match	83.2%;	Score	20.8;	DB	9;	Length	878;
Best Local Similarity	91.7%;	Pred. No.	62;				
Matches	22;	Conservative	0;	Mismatches	2;	Indels	0;
						Gaps	0;
QY	2	AAAAACACTTGGAGGGGAGCTCA	25				
Db	635	AAAGACACTTGTAAAGGGAGCTCA	658				
RESULT 8							
AY493302/c							
LOCUS							
DEFINITION	Anas platyrhynchos microsatellite CAUD057 sequence.	412 bp	DNA	linear			VRT 30-NOV-2004
ACCESSION	AY493302						
VERSION	AY493302.1	GI:40794613					
KEYWORDS							
SOURCE	Anas platyrhynchos						
ORGANISM	Anas platyrhynchos						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.						
REFERENCE	1 (bases 1 to 412)						
AUTHORS	Huang,Y.H., Tu,J.F., Cheng,X.B., Tang,B., Hu,X.X., Liu,Z.L., Feng,J.D., Lou,Y.K., Lin,L., Xu,K., Zhao,Y.D. and Li,N.						
TITLE	Isolation and characterization of 102 novel microsatellite DNA markers from the duck (Anas platyrhynchos) genome						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 412)						
AUTHORS	Huang,Y.H., Tu,J.F., Cheng,X.B., Tang,B., Hu,X.X., Liu,Z.L., Feng,J.D., Lou,Y.K., Lin,L., Xu,K., Zhao,Y.D. and Li,N.						
TITLE	Direct Submission						
JOURNAL	Submitted (01-DEC-2003) State Key Laboratory for Agrobiotechnology, China Agricultural University, Yuanmingyuan West Road 2, Beijing 100094, China						
FEATURES	Location/Qualifiers 1..412 /organism="Anas platyrhynchos" /mol_type="genomic DNA" /db_xref="taxon:8839"						
repeat_region	1..412 /note="microsatellite CAUD057" /rpt_type=tandem						
ORIGIN							
Query Match	80.8%;	Score	20.2;	DB	5;	Length	412;
Best Local Similarity	88.0%;	Pred. No.	1.3e+02;				
Matches	22;	Conservative	0;	Mismatches	3;	Indels	0;
						Gaps	0;
QY	1	AAAAACACTTGGAGGGGAGCTCA	25				
Db	304	AAAAACAGATGAAGGGAGCTCA	280				
RESULT 9							
AC165233/c							
LOCUS							
DEFINITION	Mus musculus chromosome 8 clone RP23-30702 map 8, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	158857 bp	DNA	linear			HTG 26-JUL-2005
ACCESSION	AC165233						
VERSION	AC165233.4	GI:71143318					
KEYWORDS	HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.						
SOURCE	Mus musculus (house mouse)						

ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.	
	REFERENCE	1 (bases 1 to 158857)
	AUTHORS	Birren,B., Nusbaum,C. and Lander,E.
	TITLE	Mus musculus chromosome 8, clone RP23-30702
	REFERENCE	Unpublished
	AUTHORS	2 (bases 1 to 158857)
	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collumore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	
	Direct Submission	
	Submitted (07-JUL-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA	
TITLE	JOURNAL	3 : (bases 1 to 158857)
	REFERENCE	Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collumore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
	Direct Submission	
	Submitted (26-JUL-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA	
	On Jul 26, 2005 this sequence version replaced gi:71061547.	
	All repeats were identified using RepeatMasker:	
	http://ftp.genome.washington.edu/RM/RepeatMasker.html	
	Smit, A.F.A. & Green, P. (1996-1997)	
	----- Genome Center	
	Center: Broad Institute of MIT and Harvard	
TITLE	JOURNAL	Center code: WBIR
	REFERENCE	Web site: http://www-seq.wi.mit.edu
	AUTHORS	Contact: sequence_submissions@broad.mit.edu
	----- Project Information	
	Center project name: L33365	
	Center clone name: 307_O_2	
	-----	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 3 contigs. The true order of the pieces	
FEATURES	source	
	1..158857	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10090"	
	/chromosome="8"	
	/map="8"	
	/clone="RP23-30702"	
	/clone_lib="RPCI-23 Female Mouse BAC"	
	44785..44884	
ORIGIN	gap	
	44785..44884	
	/estimated_length=unknown	
	78012..78111	
	/estimated_length=unknown	
	Query Match 80.8%; Score 20.2; DB 14; Length 158857;	
	Best Local Similarity 88.0%; Pred. No. 59;	
	Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	QY 1 AAAAAACACTTGGAGGGAGACTCA 25	
	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	
DB	80078 AAAAAACACTTGGAGGGAGAAATCA 80054	
	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	
	RESULT 10	
	AC142350/c	
	LOCUS	
	Pan troglodytes BAC clone RP43-12B4 from Y, complete sequence.	
	AC142350	
	AC142350.1 GI:29294469	
	HTG.	
	SOURCE	
REFERENCE	Pan troglodytes (chimpanzee)	
	Pan troglodytes	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.	
	Isak,A. and Haakenson,W.	
	1 (bases 1 to 168803)	
	The sequence of Pan troglodytes BAC clone RP43-12B4	
	Unpublished (2001)	
	2 (bases 1 to 168803)	
	Sulston,J.E. and Wilson,R.	
	Sequencing of Pan troglodytes	
REFERENCE	Unpublished (2001)	
	3 (bases 1 to 168803)	
	Wilson,R.K.	
	Direct Submission	
	Submitted (27-MAR-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
	4 (bases 1 to 168803)	
	Wilson,R.	
	Direct Submission	
	Submitted (18-SEP-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	
	----- Genome Center	
COMMENT	Center: Washington University Genome Sequencing Center	
	Center code: WUGSC	
	Web site: http://genome.wustl.edu	
	Contact: submissions@wustl.edu	
	----- Summary Statistics	
	Center project name: C_PT012804	
	-----	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 3 contigs. The true order of the pieces	



NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Chimpanzee Chromosome Y Mapping Project (Jennifer F. Hughes, Tatyana Pyntikova, Helen Skaletsky, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RP43 BAC library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (<http://www.resgen.com>) or Pieter de Jong and co-workers at <http://www.bacpac.chori.org>.

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

Location/Qualifiers

1..168803  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="Y"  
/map="Y"  
/clone="RP43-12B4"  
/clone\_lib="RP43-43"

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 168803;  
Best Local Similarity 88.0%; Pred. No. 58;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACTCA 25

Db 99854 AAAAAACATGGAGAGATCACTCA 99830

RESULT 11  
AC147666/c 189158 bp DNA linear PRI 10-JUN-2004  
LOCUS Pan troglodytes BAC clone CH251-19607 from Y, complete sequence.  
DEFINITION  
AC147666  
ACCESSION  
VERSION AC147666.4 GI:49458043  
KEYWORDS HTG.

SOURCE Pan troglodytes (chimpanzee)

#### ORGANISM

Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.

REFERENCE 1 (bases 1 to 189158)

AUTHORS Belter, E., Haglund, K., Elliott, G. and Meyer, R.

TITLE The sequence of Pan troglodytes BAC clone CH251-19607

JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 189158)

AUTHORS Wilson, R.K.

TITLE Direct Submission

JOURNAL Submitted (21-DEC-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

#### REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

#### COMMENT

3 (bases 1 to 189158)

Wilson, R.K.

Direct Submission

Submitted (24-APR-2004) Genetics, Genome Sequencing Center, 4444

Forest Park Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 189158)

Wilson, R.K.

Direct Submission

Submitted (30-JUN-2004) Washington University School of Medicine,

Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO

63108, USA

On Jun 30, 2004 this sequence version replaced gi:46395453.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

----- Summary Statistics

----- Center project name: C\_AB0196007

-----

#### NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Chimpanzee Chromosome Y Mapping Project (Jennifer F. Hughes, Tatyana Pyntikova, Helen Skaletsky, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The CHORI-251 Chimpanzee BAC library has been constructed at the Children's Hospital Oakland Research Institute, BACPAC Resources, by Dr. Baoli Zhu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from Pieter de Jong and co-workers at <http://www.bacpac.chori.org>.

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC147127 and AC146455.

#### FEATURES

source

Location/Qualifiers

1..189158  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="Y"  
/map="Y"

misc\_feature

/clone="CH251-19607"

/clone\_lib="CHORI251"

29362..29508

/note="Sequence derived from PCR product of project DNA."

unsure

30013..30125

/note="Sequence derived from one plasmid subclone."

misc\_feature

30471..30784

/note="Sequence derived from PCR product of project DNA."

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 189158;

Best Local Similarity 88.0%; Pred. No. 58;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACTCA 25

```

Db      170949 AAAAAACATCGAAGAGACTCA 170925
|||||
AC145502      196620 bp      DNA      linear      HTG 08-AUG-2003
LOCUS      Canis familiaris clone RP81-204N10, WORKING DRAFT SEQUENCE.
AC145502
ACCESSION
VERSION      AC145502.1 GI:32964888
KEYWORDS      HTG; HTGS-PHASE2; HTGS DRAFT.
SOURCE      Canis familiaris (dog)
ORGANISM      Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE      1 (bases 1 to 196620)
AUTHORS      Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Carliaga,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Lalic,P., Lee-Lin,S.-O.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C.,
Maskeri,B., McDowell,J., Paquirigan,C., Pearson,R., Portnoy,M.E.,
Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
TITLE      NISC Comparative Sequencing Initiative
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 196620)
AUTHORS      Green,E.D.
TITLE      Direct Submission
JOURNAL      Submitted (18-JUL-2003) NTH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
REFERENCE      3 (bases 1 to 196620)
AUTHORS      Green,E.D.
TITLE      Direct Submission
JOURNAL      Submitted (08-AUG-2003) NTH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
COMMENT      ----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hri.nih.gov
----- Project Information
Center project name: etm
Center clone name: 204N10

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 196374 bases at least Q40
Consensus quality: 196562 bases at least Q30
Consensus quality: 196601 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 196620; sum-of-contigs
Quality coverage: 10.83x in Q20 bases; agarose-fp
Quality coverage: 10.19x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* the accession number will be 196620 bp in length.
FEATURES             source
Location/Qualifiers
1..196620
/organism="Canis familiaris"
/mol_type="genomic DNA"
/db_xref="taxon:9615"
/clone="RP81-204N10"
/clone_lib="RP81"
1..196620
/notes="assembly_fragment"
clone_end:77
vector_side:left
clone_end:Sp6
vector_side:right
ORIGIN
Query Match      80.8%; Score 20.2; DB 14; Length 196620;
Best Local Similarity 88.0%; Pred. No. 57;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      1 AAAAAACACTTGAAGGAGACTCA 25
Db      185805 ACAACACACTTGCAGGAGACTCA 185829
|||||
AC102363      210317 bp      DNA      linear      ROD 22-JAN-2005
LOCUS      Mus musculus chromosome 1, clone RP23-76B18, complete sequence.
AC102363
ACCESSION      AC102363.19 GI:58036679
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 210317)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Mus musculus chromosome 1, clone RP23-76B18
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 210317)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gaidyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Haeft,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lakoczek,K.,
Lamaze,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigliio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission

```

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 210317)  
 Birren,B., Nubaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagoian,D., Hagos,B., Hall,J.J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (19-OCT-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 210317)

## REFERENCE

## AUTHORS

Birren,B., Nubaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagoian,D., Hagos,B., Hall,J.J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (22-JAN-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Jan 22, 2005 this sequence version replaced gi:54291951.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information

Center project name: L18576

Center clone name: 76\_B\_18

-----

## FEATURES

## source

Location/Qualifiers

1..210317

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="1"

/map="1"

/clone="RP23-76B18"

/clone\_lib="RPCI-23 Female Mouse BAC"  
 complement(55..1188)  
 /rpt\_family="Lx4"  
 1307..1368  
 /rpt\_family="GA)n"  
 1465..1645  
 /rpt\_family="Lx8"  
 1855..2172  
 /rpt\_family="Lx8"  
 complement(2450..2823)  
 /rpt\_family="MTA"  
 3335..4262  
 /rpt\_family="Lx2B"  
 4263..4287  
 /rpt\_family="AT\_rich"  
 4385..4460  
 /rpt\_family="(TA)n"  
 complement(5005..5347)  
 /rpt\_family="ORR1B1"  
 5348..5376  
 /rpt\_family="(T)n"  
 complement(5401..5717)  
 /rpt\_family="L1\_MM"  
 5769..5819  
 /rpt\_family="L1M2"  
 6130..6170  
 /rpt\_family="AT\_rich"  
 6648..6826  
 /rpt\_family="B3"  
 6901..7087  
 /rpt\_family="Lx"  
 7098..7141  
 /rpt\_family="(TAA)n"  
 7142..7166  
 /rpt\_family="(TTA)n"  
 7168..9087  
 /rpt\_family="Lx"  
 10295..10670  
 /rpt\_family="ORR1B2"  
 11203..11253  
 /rpt\_family="(TAAA)n"  
 11297..11394  
 /rpt\_family="AT\_rich"  
 complement(12397..12888)  
 /rpt\_family="L1\_MM"  
 13083..13745  
 /rpt\_family="RMER6A"  
 14619..15735  
 /rpt\_family="L1VL2"  
 15736..15821  
 /rpt\_family="(TA)n"  
 15822..19354  
 /rpt\_family="L1VL2"  
 19342..19713  
 /rpt\_family="L1\_MM"  
 19708..21263  
 /rpt\_family="Lx6"  
 21256..22097  
 /rpt\_family="Lx"  
 complement(22304..22464)  
 /rpt\_family="MT2A"  
 23133..23184  
 /rpt\_family="(TA)n"  
 23826..23854  
 /rpt\_family="(A)n"  
 24847..24980  
 /rpt\_family="B4"  
 25004..25038  
 /rpt\_family="(TCTG)n"  
 25040..25064  
 /rpt\_family="(TA)n"  
 25370..25395  
 /rpt\_family="(TTTG)n"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

HTG 23-JAN-2005

DEFINITION	Bos taurus clone CH240-4L7, WORKING DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION	AC152260
VERSION	AC152260.2 GI:58038041
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Anarstange,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buay,C., Burch,P., Burnett,C., Burrell,K.L., Byrd,N.C., Chen,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Garisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrill,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Huly,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,B., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Nguyen,N., Nickerson,B., Nwokenkwo,S., Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaik,I., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Verduzco,D., Villalon,D., Vinson,R., Waldron,L.M., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE	Direct Submission
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 215892)
TITLE	Worley,K.C.
JOURNAL	Direct Submission
AUTHORS	Submitted (01-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 215892)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (23-JAN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jan 23, 2005 this sequence version replaced gi:55058582. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FAFR  
Center clone name: CH240-4L7  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 209512 bases at least Q40  
Consensus quality: 211147 bases at least Q30  
Consensus quality: 212594 bases at least Q20  
Estimated insert size: 211936; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1232: contig of 1232 bp in length  
1233 1282: gap of 50 bp  
1283 20143: contig of 18861 bp in length  
20144 20193: gap of 50 bp  
20194 28382: contig of 8189 bp in length  
28383 28432: gap of 50 bp  
28433 50691: contig of 22259 bp in length  
50692 50791: gap of unknown length  
50792 85570: contig of 34779 bp in length  
85571 85620: gap of 50 bp  
85621 90790: contig of 5170 bp in length  
90791 90929: gap of 139 bp  
90930 95104: contig of 4175 bp in length  
95105 95154: gap of 50 bp  
95155 102719: contig of 7565 bp in length  
102720 102769: gap of 50 bp  
102770 134604: contig of 31835 bp in length  
134605 134704: gap of unknown length  
134705 137956: contig of 3252 bp in length  
137957 138006: gap of 50 bp  
138007 153576: contig of 15570 bp in length  
153577 153626: gap of 50 bp  
153627 162660: contig of 9034 bp in length  
162661 163236: gap of 576 bp  
163237 198080: contig of 34844 bp in length  
198081 198405: gap of 325 bp  
198406 200117: contig of 1712 bp in length  
200118 207948: contig of 50 bp  
207949 207998: gap of 50 bp  
207999 209885: contig of 1887 bp in length  
209886 209985: gap of unknown length  
209986 211031: contig of 1046 bp in length  
211032 211132: gap of unknown length  
211133 212180: contig of 1049 bp in length  
212181 212280: gap of unknown length  
212281 213280: contig of 1048 bp in length  
213282 213428: gap of unknown length  
213429 215892: contig of 2464 bp in length.  
213429 Location/Qualifiers  
1. 215892  
/organism="Bos taurus"

```

/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-4L7"
1233..1282
/estimated_length=50
20144..20193
/estimated_length=50
28383..28432
/estimated_length=50
50892..50791
/estimated_length=unknown
85571..85620
/estimated_length=50
90791..90929
/estimated_length=139
95105..95154
/estimated_length=50
102720..102769
/estimated_length=50
134605..134704
/estimated_length=unknown
137957..138006
/estimated_length=50
153577..153626
/estimated_length=50
162661..163236
/estimated_length=576
198081..198405
/estimated_length=325
200118..200167
/estimated_length=50
207949..207998
/estimated_length=50
209886..209985
/estimated_length=unknown
211032..211131
/estimated_length=unknown
212181..212280
/estimated_length=unknown
213329..213428
/estimated_length=unknown

```

## ORIGIN

```

Query Match      80.8%; Score 20.2; DB 14; Length 215892;
Best Local Similarity 88.0%; Pred. No. 57;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAACACTTGGAGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
Db 108402 AAAAAACCTAGAGGAGACTCA 108378

```

```

RESULT 16
AC116092 Rattus norvegicus clone CH230-134D13, *** SEQUENCING IN PROGRESS
LOCUS AC116092 230954 bp DNA linear HTG 08-OCT-2002
DEFINITION Rattus norvegicus clone CH230-134D13, *** SEQUENCING IN PROGRESS
AC116092
AC116092.3 GI:23269875
VERSION HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus

```

```

ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 230954)
Murny D.Marie., Metzker M.Lee., Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,

```

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karkathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhera, L., Loulsegged, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norrie, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villabana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.

## TITLE

Unpublished

2 (bases 1 to 230954)

Worley, K. C.

Direct Submission

Submitted (24-MAR-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 230954)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 23, 2002 this sequence version replaced gi:21745909.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>



```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRLW
Center clone name: CH230-134D13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 174870 bases at least Q40
Consensus quality: 178840 bases at least Q30
Consensus quality: 181745 bases at least Q20
Estimated insert size: 199572; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4497: contig of 4497 bp in length
* 4498 4597: gap of unknown length
* 4598 25299: contig of 20702 bp in length
* 25300 25399: gap of unknown length
* 25400 208371: contig of 182972 bp in length
* 208372 208471: gap of unknown length
* 208472 209709: contig of 1238 bp in length
* 209710 209809: gap of unknown length
* 209810 210979: contig of 1170 bp in length
* 210980 211079: gap of unknown length
* 211080 212892: contig of 1813 bp in length
* 212893 212992: gap of unknown length
* 212993 218147: contig of 5155 bp in length
* 218148 218247: gap of unknown length
* 218248 219391: contig of 1144 bp in length
* 219392 219491: gap of unknown length
* 219492 225125: contig of 5634 bp in length
* 225126 225225: gap of unknown length
* 225226 227317: contig of 2092 bp in length
* 227318 227417: gap of unknown length
* 227418 230954: contig of 3537 bp in length.
*
* Location/Qualifiers
* source
* 1. .230954
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-134D13"
*
* misc_feature
* 1. .1265
* /note="wgs contig"
* 4498. .4597
* /estimated_length=unknown
* 21796. .22828
* /note="wgs contig"
* 25300. .25399
* /estimated_length=unknown
* 138428. .139768
* /note="wgs contig"
* 208372. .208471
* /estimated_length=unknown
* 209710. .209809
* /estimated_length=unknown
* 210980. .211079
* /estimated_length=unknown
* 212893. .212992
* /estimated_length=unknown
* 218148. .218247
* /estimated_length=unknown
* 219392. .219491
* /estimated_length=unknown
* 225126. .225225
* /estimated_length=unknown

```

```

gap 227318. .227417
/estimated_length=unknown

ORIGIN
Query Match 80.8%; Score 20.2; DB 14; Length 230954;
Best Local Similarity 88.0%; Pred. No. 56;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACTCA 25
|||||
Db 110889 AAAAAACACTTGGAGGGAGACTGA 110913

RESULT 17
LOCUS AC140677 121010 bp DNA linear PRI 27-FEB-2003
DEFINITION Homo sapiens chromosome 5 clone CTC-778L3, complete sequence.
ACCESSION AC140677
VERSION AC140677.1 GI:28570298
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 121010)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 121010)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
1. .121010
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="5"
/clone="CTC-778L3"

ORIGIN
Query Match 79.2%; Score 19.8; DB 8; Length 121010;
Best Local Similarity 91.3%; Pred. No. 95;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACT 23
|||||
Db 6013 AAAAAAATTTGGAGGGAGACT 6035

RESULT 18
LOCUS AC114959 163805 bp DNA linear PRI 22-MAY-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-170L13, complete sequence.
ACCESSION AC114959
VERSION AC114959.2 GI:21070666
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 163805)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```



```

JOURNAL
REFERENCE
TITLE
AUTHORS
JOURNAL
JOURNAL
REFERENCE
TITLE
AUTHORS
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Query Match
Best Local Similarity
Matches
21; Conservative
0; Mismatches
2; Indels
0; Gaps
0;
Qy
1
AAAAAAAACTGGAGGAGACT 23
|||||
Db
5076
AAAAAAAAATTGGAAGGGAGACT 5098
|||||
RESULT 19
AC156954/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 163805)
DOE Joint Genome Institute.
Direct Submission
Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163805)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (22-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On May 22, 2002 this sequence version replaced gi:19424420.
Draft Sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. 163805
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-170L13"
79.2%; Score 19.8; DB 8; Length 163805;
Pred. No. 92;
71.3%;
21; Conservative
0; Mismatches
2; Indels
0; Gaps
0;
AC156954
Mus musculus 6 BAC RP24-343M7 DNA linear ROD 30-MAY-2005
(C57BL/6J Male) Mouse BAC Library) complete sequence.
AC156954 AC116805
AC156954.5 GI:66793675
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 179140)
Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadore,J., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaslin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haerberlen,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hognes,M., Hollins,B.,
Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,P.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
London,P., Lopez,J., Lorensuhewa,L., Lozado,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Muntadaa,M.,
Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
Nott,A., Nwaokemele,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,
Okwono,G., Okwono,K., Parker,D., Pasternak,S., Patel,B.,
Patei,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
Rabata,C., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,P., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,B., Song,X.-Z., Sorrelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
Trejos,Z., Usmami,K., Vargo,C., Verduzco,D., Villaseana,D., Virk,D.,
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 179140)
Worley,K.C.
Direct Submission
Submitted (06-FEB-2005) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 179140)
Worley,K.C.
Direct Submission
Submitted (30-MAY-2005) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 179140)
Worley,K.C.
Direct Submission
Submitted (30-MAY-2005) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 30, 2005 this sequence version replaced gi:62122979.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using PowerBlast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: WIBR
Contact: hgsc-help@bcm.tmc.edu.
Location/Qualifiers
1. 179140
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/clone="RP24-343M7"
1. 86096
/notes="overlaps bases 133227..219322 of clone AC153628"
/function="clone overlap"
9. 118
/rpt_family="WTD"
119. 199
/rpt_family="(GGA)A)n"
200..260
/rpt_family="WTD"
FEATURES
source
misc_feature
repeat_region
repeat_region
repeat_region

```

```
repeat_region complement(956..1176)
/rpt_family="MTD"
1177..1222
/rpt_family="(TGGA)n"
repeat_region complement(1223..1412)
/rpt_family="MTD"
1919..2136
/rpt_family="BAA"
repeat_region 2137..2164
/rpt_family="(CAAAA)n"
repeat_region 2180..2330
/rpt_family="GA-rich"
repeat_region 2510..2549
/rpt_family="(GAAAA)n"
repeat_region 2637..2692
/rpt_family="(CA)n"
repeat_region 2693..2751
/rpt_family="(GA)n"
repeat_region 2771..2847
/rpt_family="ID"
repeat_region 2848..2876
/rpt_family="(CAAT)n"
repeat_region 2971..3022
/rpt_family="MTD-int"
repeat_region 3023..3068
/rpt_family="(TA)n"
repeat_region 3070..3094
/rpt_family="(GA)n"
repeat_region 3095..3143
/rpt_family="MTD-int"
repeat_region 3145..3696
/rpt_family="Lx7"
repeat_region 3697..3716
/rpt_family="(TTGG)n"
repeat_region 3717..3867
/rpt_family="Lx7"
repeat_region 3949..4205
/rpt_family="MTD-int"
repeat_region 4213..4349
/rpt_family="B1_Mm"
repeat_region 4351..4436
/rpt_family="A-rich"
repeat_region 4585..5259
/rpt_family="MTD-int"
repeat_region 5277..6562
/rpt_family="L1Md_F"
repeat_region complement(6563..6668)
/rpt_family="L1Md_F"
repeat_region complement(6679..6890)
/rpt_family="ORR1D-int"
repeat_region 6891..9247
/rpt_family="L1_Mm"
repeat_region 9248..9310
/rpt_family="L1_Mus1"
repeat_region 9312..9332
/rpt_family="AT-rich"
repeat_region 9335..9435
/rpt_family="MTD-int"
repeat_region 9450..9771
/rpt_family="MTE"
repeat_region complement(9808..10047)
/rpt_family="Arthurl1"
repeat_region 11734..11838
/rpt_family="(CAGA)n"
repeat_region complement(11936..12062)
/rpt_family="RSINEL"
repeat_region complement(12120..12278)
/rpt_family="MRB"
repeat_region complement(12663..12839)
/rpt_family="B4A"
repeat_region complement(12865..13472)
/rpt_family="Lx3B"
repeat_region complement(14128..14290)
```

---

```
repeat_region /rpt_family="MERSB"
15935..15956
/rpt_family="AT-rich"
repeat_region 17347..17392
/rpt_family="(TG)n"
repeat_region complement(17607..18001)
/rpt_family="MLTID"
18371..18652
/rpt_family="L2"
repeat_region complement(19155..19548)
/rpt_family="ORR1C1"
20910..21032
/rpt_family="B1_Mus1"
repeat_region complement(21556..22056)
/rpt_family="L1Md_Gf"
repeat_region 22057..22583
/rpt_family="L1Md_T"
repeat_region 22617..22640
/rpt_family="(A)n"

Query Match 79.2%; Score 19.8; DB 9; Length 179140;
Best Local Similarity 91.3%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAACACTTGAAGGAGACTCA 25
|||||
Db 74776 AAAGTCTTGAAGGAGACTCA 74754

RESULT 20
AC131565/c
LOCUS AC131565 181283 bp DNA linear PRI 27-AUG-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-1273L8, complete sequence.
ACCESSION AC131565
VERSION AC131565.1 GI:22474846
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 181283)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181283)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 181283)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT www.jgi.doe.gov
Draft Sequence Produced by DOE Joint Genome Institute
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
source
1..181283
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1273L8"

ORIGIN

Query Match 79.2%; Score 19.8; DB 8; Length 181283;
Best Local Similarity 91.3%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY

1 AAAAAACACTTGAAGGAGACT 23

Db

177390 AAAAAAATTTGAAGGAGACT 177368

RESULT 21

AF129075/c

LOCUS

DEFINITION

AF129075 190837 bp DNA linear PRI 13-AUG-2003  
Homo sapiens chromosome 21 clone RP1-100J12, RP1-P79E4 map  
q21.3-22.11, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AF129075 3 GI:33620784  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 190837)  
Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T.,  
Park, H.-S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.-K., Soeda, E.,  
Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,  
Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,  
Patterson, D., Reichwald, K., Rump, A., Schillhabel, M. B., Schudy, A.,  
Zimmermann, W., Rosenthal, A., Kudoh, J., Kawasaki, K., Asakawa, S.,  
Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S.,  
Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordieck, G.,  
Hornischer, K., Brandt, P., Scharte, M., Schoen, O., Desario, A.,  
Reichelt, J., Kauer, G., Bloecker, H., Ramser, J., Beck, A., Klages, S.,  
Hennig, S., Rieselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S.,  
Borzym, K., Gardiner, K., Mizetic, D., Francis, P., Lehrach, H.,  
Reinhardt, R., and Yaspo, M. Laure.

The DNA sequence of human chromosome 21

Unpublished

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 190837)  
Taudien, S., Dagand, E., Delabar, J., Orti, R., Nordieck, G.,  
Drescher, B., Weber, J., Schattevoy, R., Menzel, U., Yaspo, M.-L. and  
Rosenthal, A.

TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission  
Submitted (16-FEB-1999) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
3 (bases 1 to 190837)  
Taudien, S., Dagand, E., Delabar, J., Orti, R., Nordieck, G.,  
Drescher, B., Weber, J., Schattevoy, R., Menzel, U., Yaspo, M.-L. and  
Rosenthal, A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

COMMENT

Submitted (20-APR-2001) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
4 (bases 1 to 190837)  
Legemann, D. and Platzer, M.  
Direct Submission  
Submitted (13-AUG-2003) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
On Aug 13, 2003 this sequence version replaced gi:13699324.

----- Genome Center  
Center: Institute of Molecular Biotechnology  
Center code: IMB  
Web site: <http://genome.imb-jena.de/>  
Contact: [gscj-submit@genome.imb-jena.de](mailto:gscj-submit@genome.imb-jena.de)  
----- Project Information  
Center project name: U70-64  
Center clone name: RP1-100J12, RP1-P79E4  
----- Consensus quality: 188276 bases at least Q40  
Consensus quality: 130662 bases at least Q30  
Consensus quality: 190837 bases at least Q20  
Quality coverage: 10.52x  
-----

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest.

-----  
Neighboring sequence information:

This clone is overlapped by RP5-866H8, RP1-84N21.

## FEATURES

## source

## Location/Qualifiers

1. .190837

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="q21.3-22.11"

/clone="RP1-100J12, RP1-P79E4"

1. .121509

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="RP1-100J12"

48002. .190837

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="RP1-79E4"

653

/note="C substituted in clone: RP5-866H8"

/replaces="t"

1060

/note="A substituted in clone: RP5-866H8"

/replaces="g"

1314

/note="C substituted in clone: RP5-866H8"

/replaces="t"

3311

/note="C substituted in clone: RP5-866H8"

/replaces="t"

3725

/note="G substituted in clone: RP5-866H8"

/replaces="a"

3827

/note="deleted in clone: RP1-100J12"

/replaces=""

5555

/note="T substituted in clone: RP5-866H8"

/replaces="c"

6971

/note="deleted in clone: RP5-866H8"

/replaces=""

7254

/note="G substituted in clone: RP5-866H8"

/replaces="c"

8668

/note="G substituted in clone: RP5-866H8"

/replaces="a"

15237. .15255

/note="single stranded/single chemistry region"

26658. .26695

/note="single stranded/single chemistry region"

30914. .31021

/note="single stranded/single chemistry region"

30916. .30919

/note="low quality region"

30919

39287. .39356

/note="single stranded/single chemistry region"

39705. .39766

/note="single stranded/single chemistry region"

44251

/note="low quality region"

44287. .44367

/note="single stranded/single chemistry region"

48080

/note="C substituted in clone: RP1-79E4"

```

variation      /replace="g"
49109          /note="G substituted in clone: RP1-79E4"
/replace="t"
variation      49195
/replace="t"
variation      49693
/replace="a"
/replace="a"
variation      52403
/replace="a"
/replace="a"
variation      53247
/replace="a"
/replace="a"
variation      53545
/replace="a"
/replace="a"
variation      53696
/replace="a"
/replace="a"
variation      58070
/replace="a"
/replace="a"
variation      60115
/replace="a"
/replace="a"
misc_feature   61768
/replace="a"
/replace="a"
variation      63716
/replace="a"
/replace="a"
variation      70290
/replace="a"
/replace="a"
misc_feature   73369
/replace="a"
/replace="a"
misc_feature   73856
/replace="a"
/replace="a"
misc_feature   74003
/replace="a"
/replace="a"
variation      74435
/replace="a"
/replace="a"
variation      74449
/replace="a"
/replace="a"
variation      74490
/replace="a"
/replace="a"
variation      75076
/replace="a"
/replace="a"
variation      75753
/replace="a"
/replace="a"
variation      78907
/replace="a"
/replace="a"
misc_feature   79856
/replace="a"
/replace="a"
/notes="single stranded/single chemistry region"
79.2%; Score 19.8; DB 8; Length 190837;
Query Match    91.3%; Pred. No. 90;
Best Local Similarity
Matches 21; Conservative 0; Mismatches 2; Indels 0; Caps 0;
OY 1 AAAAAACACTTGGAGGAGACT 23
||||| ||||| ||||| ||||| |||||
Db 113272 AAAAAAACTGTGAAGGAGACT 113250

```

```

BS000157/c
LOCUS          192836 bp      DNA      linear      PRI 09-JUN-2004
DEFINITION    Pan troglodytes chromosome 22 clone:RP43-154K22, map 22, complete
sequences.
ACCESSION     BS000157
VERSION       BS000157.1
KEYWORDS      HTG.
SOURCE        Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
1
The International Chimpanzee Chromosome 22 Consortium.
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429, 382-388 (2004)
2 (bases 1 to 192836)
Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
Fujiyama,A. and Sakaki,Y.
Direct Submission
Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBP, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
----- Genomic Center
Center: RIKEN Genomic Sciences Center
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project
Center Clone name: RP43-154K22
----- Summary Statistics
Sequencing vector: pUC18,pUC13,prTZ19B; 100% of reads Chemistry:
Dye-terminator Big Dye and ET; 100% of reads Assembly Program:
Phrap; version 0.990329
Consensus quality: 191,815 bases at least Q40
Consensus quality: 960 bases at least Q30
Consensus quality: 61 bases at least Q20
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.
-----
Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pBACE3.6

```

The CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Cline') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoji Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Bichler & Pieter J de Jong. The library characteristics are described at <http://www.chori.org/bacpac/chimpanzee251.htm>. The clone may be obtained from Pieter J. de Jong and coworkers (<http://www.chori.org/bacpac>).

VECTOR: pTARBAC2.1  
The PTB1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.

Clones may be obtained from Asao Fujiyama and co-workers (<http://www.gsc.riken.go.jp>).

VECTOR: pKS145

The PT722 chimpanzee Fosmid library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.

Clones may be obtained from Asao Fujiyama and co-workers (<http://www.gsc.riken.go.jp>).

VECTOR: pKS143

#### Sequence Quality Assessment:

This entry has been annotated with sequence estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Neighboring clones: PTB-083K24 (left) and PTB-086M04 (right).

#### FEATURES

source

Location/Qualifiers

1..192836  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="22"  
/clone="RP43-154K22"  
/clone\_lib="RPCI-43 chimpanzee BAC"

#### ORIGIN

Query Match 79.2%; Score 19.8; DB 8; Length 192836;  
Best Local Similarity 91.3%; Pred. No. 89;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGGAAGGAGACT 23

Db 25268 AAAAAAACTTGTAAAGGAGACT 25246

RESULT 23

AL954132/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

1

Submitted (05-JUN-2003)

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk

On Jun 6, 2003 this sequence version replaced gi:31407707.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>  
Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

[http://www.sanger.ac.uk/Projects/D\\_rexio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rexio/fishmask.shtml) DKEY-32N7 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5

Location/Qualifiers

1..197745  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-32N7"  
/clone\_lib="DanioKey"

#### ORIGIN

Query Match 79.2%; Score 19.8; DB 5; Length 197745;  
Best Local Similarity 91.3%; Pred. No. 89;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGGGAAGGAGACT 24

Db 62477 AAAAAACACTTGGGAAGGAGACT 62455

RESULT 24

BS000171/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

1

The International Chimpanzee Chromosome 22 Consortium.

DNA sequence and comparative analysis of chimpanzee chromosome 22

Nature 429, 382-388 (2004)

2 (bases 1 to 218680)

Wang, S., Cai, Z., Wang, B., Zheng, H., Zhang, Y., Zhang, X., Zhu, G., Lu, G., Fu, G. and Chen, Z.

Direct Submission

Submitted (26-MAY-2003)

Shengyue Wang, Chinese National Human Genome Center at Shanghai, Shanghai 201203, CHINA

(E-mail: [wangyue@chgc.sh.cn](mailto:wangyue@chgc.sh.cn), URL: <http://www.chgc.sh.cn>, Tel: 86-21-50801919, Fax: 86-21-50801922)

COMMENT	The Chimpanzee Chromosome 22 Sequencing Consortium consists of: *Chinese National Human Genome Center at Shanghai, Shanghai, China; *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research Center, Daejeon, Korea; *Max-Planck-Institute for Molecular Genetics, Berlin, Germany; *National Institute of Genetics, Mishima, Japan; *National Yang Ming University Genome Research Center, Taipei, Taiwan; *RIKEN Genomic Sciences Center, Yokohama, Japan. ----- Genomic Center Center: Chinese National Human Genome Center at Shanghai Center code: CHGCS Web site: <a href="http://chgc.sh.cn">http://chgc.sh.cn</a> Contact: <a href="mailto:wangsy@chgc.sh.cn">wangsy@chgc.sh.cn</a> ----- Project Information Center project name: The Chimpanzee Chromosome 22 Sequencing Project Center clone name: PTB-086M04 ----- Summary Statistics Sequencing vector: pUC18, 100% of reads Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 218104 bases at least Q40 Consensus quality: 218625 bases at least Q30 Consensus quality: 218668 bases at least Q20 Quality coverage: 9.3x ----- This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. ----- Source information: The PTB1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee. Clones may be obtained from Asao Fujiyama and co-workers ( <a href="http://www.gsc.riken.go.jp">http://www.gsc.riken.go.jp</a> ). VECTOR: pKS145 Sequence Quality Assessment: This entry has been annotated with sequence estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. ----- Neighboring clones: PTB-083K24 (left) and PTB-016G04 (right). -----		
FEATURES	source	Location/Qualifiers	
		1. 218680 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" /chromosomes="22" /clone="PTB-086M04" /clone_lib="PTB1 chimpanzee BAC"	
ORIGIN			
Query Match	79.2%;	Score 19.8; DB 8; Length 218680;	
Best Local Similarity	91.3%;	Pred. No. 88;	
Matches	21; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
Oy	1	AAAAACACTTGAAGGGAGACT 23 	
Db	16572	AAAAAAACTTGTAAAGGGAGACT 16550	
RESULT 25			
AC153628	219322 bp	DNA linear	ROD 30-MAY-2005
LOCUS	Mus musculus 6 BAC RP23-13B24 (Roswell Park Cancer Institute (C57BL/6J Female) Mouse BAC Library) complete sequence.		
DEFINITION	AC153628 AC113959		
ACCESSION	AC153628.3 GI:58330944		
VERSION	HTG.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 219322)		
AUTHORS	Murny, D., Adams, C., Agbai II, O., Allen, C., Albrooks, S., Archer, P., Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R., Beraducci, A., Biswal, K., Blyth, P., Bonham H., Buhay, C., Burch, P., Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I., Chacko, J., Chahrouh, M., Chavez, D., Chen, A., Chen, G., Chen, R., Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A., Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K., Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando P., Flagg N., Forbes, L., Foster, P., Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T., Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haeblerlein, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hogue, M., Hollins, B., Howell, L.T., Hulyk, S., Hume, J., Ito, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, P., Leal, S., Lee, K., Lee, S., LeGall, F.I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorensuhewa, L., Lozada, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McLelland, H., McPherson, J., Mercadao, C., Metzker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Nott, A., Nwaokemele, O., Obregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S.J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y., Sison, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Umani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yaun, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.		
	Direct Submission		
	Unpublished		
	2 (bases 1 to 219322)		
JOURNAL	Worley, K.C.		
REFERENCE	Direct Submission		
AUTHORS	Worley, K.C.		
TITLE	Submitted (14-DEC-2004) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
JOURNAL	3 (bases 1 to 219322)		
REFERENCE	Worley, K.C.		
AUTHORS	Direct Submission		
TITLE	Submitted (18-DEC-2004) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
JOURNAL	4 (bases 1 to 219322)		
REFERENCE	Worley, K.C.		
AUTHORS	Direct Submission		
TITLE	Submitted (29-JAN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
JOURNAL	5 (bases 1 to 219322)		
REFERENCE	Worley, K.C.		
AUTHORS	Direct Submission		
TITLE			

## JOURNAL

Submitted (30-MAY-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 29, 2005 this sequence version replaced gi:56676412.  
Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

## COMMENT

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

source  
1. 219322  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="6"  
/clone="RP23-13B24"  
complement(1..142)  
/rpt\_family="L1\_Mm"  
2059..2244  
/rpt\_family="Lx4B"  
2295..2389  
/rpt\_family="GA"  
2423..2515  
/rpt\_family="GAA"  
3746..3803  
/rpt\_family="TTTG"  
3815..3846  
/rpt\_family="CCA"  
complement(4969..5013)  
/rpt\_family="L3"  
5632..5680  
/rpt\_family="TTA"  
complement(5771..6024)  
/rpt\_family="Lx8"  
7995..8304  
/rpt\_family="L1Md\_F3"  
complement(8335..8699)  
/rpt\_family="ORR1D"  
complement(8940..9174)  
/rpt\_family="L1\_Mur3"  
complement(9175..9641)  
/rpt\_family="L1\_Mus3"  
complement(9642..9674)  
/rpt\_family="L1\_Mur3"  
9900..10071  
/rpt\_family="L1Md\_F2"  
12020..12046  
/rpt\_family="TTTG"  
17140..17222  
/rpt\_family="CA"  
17277..17382  
/rpt\_family="MLTB"  
complement(17541..17588)  
/rpt\_family="L2"  
17949..18012  
/rpt\_family="CT-rich"  
18014..18124  
/rpt\_family="C-rich"  
18132..18270  
/rpt\_family="CT-rich"  
complement(19103..19270)  
/rpt\_family="L1Md\_F2"  
22045..22786  
/rpt\_family="L1Md\_F2"

repeat\_region 25160..25196  
/rpt\_family="(T)n"  
repeat\_region complement(25197..25323)  
/rpt\_family="B1\_Mus1"  
27714..27723  
/note="single subclone"  
/function="single clone coverage"  
repeat\_region 29432..29742  
/rpt\_family="L1Md\_T"  
repeat\_region complement(32227..33010)  
/rpt\_family="L1Md\_T"  
33234..33355  
/rpt\_family="(TTTC)n"  
33385..33528  
/rpt\_family="RSINE1"  
34120..34221  
/rpt\_family="RSINE1"  
34247..34274  
/rpt\_family="(TAAAA)n"  
34887..35090  
/rpt\_family="B3"  
35546..35567  
/rpt\_family="AT-rich"  
complement(36257..36351)  
/rpt\_family="MTE"  
37477..37918  
/rpt\_family="L1Md\_F3"  
complement(37939..38080)  
/rpt\_family="B1\_Mus1"  
38397..38447  
/rpt\_family="(CA)n"  
40153..40204  
/rpt\_family="(ATG)n"  
complement(40354..40521)  
/rpt\_family="URR1A"  
complement(40872..41236)  
/rpt\_family="ORR1D"  
41697..41726  
/rpt\_family="AT-rich"  
complement(42404..42502)  
/rpt\_family="L3"  
42552..42757  
/rpt\_family="B4"  
complement(42758..42903)  
/rpt\_family="B1\_Mm"  
42904..42975  
/rpt\_family="B4"  
43338..43360  
/rpt\_family="AT-rich"  
44681..44722  
/rpt\_family="(TG)n"  
45747..45804  
/rpt\_family="GA-rich"  
45989..46018  
repeat\_region

Query Match 79.2%; Score 19.8; DB 9; Length 219322;  
Best Local Similarity 91.3%; Pred. No. 88;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACACTTGGAAAGGAGACTCA 25  
|||||

Db 208002 AAAAGCTTGGAAAGGAGACTCA 207980  
|||||

## RESULT 26

HS21C049/c HS21C049 340000 bp DNA linear PRI 16-APR-2005  
LOCUS Homo sapiens chromosome 21 segment HS21C049.  
DEFINITION AL163249 AP001704 BA000005  
ACCESSION AL163249.2 GI:7717307  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 340000)

#### REFERENCE AUTHORS

Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Iehii,K., Totoki,Y., Choi,D.K., Soeda,B., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Reonthai,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.B., Minoshima,S., Shimizu,N., Nordseick,G., Hornischer,K., Brandt,P., Schafke,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Dagand,F., Wehrmeyer,S., Borzym,K., and Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R., and Yaspo,M.L.

#### TITLE JOURNAL

Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

The Chromosome 21 Mapping and Sequencing Consortium consists of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagamihara 228-8555, Japan,  
\* e.mail: sakaki@gsc.riken.go.jp  
\* URL: http://hgp.gsc.riken.go.jp/

and

\* Institute of Molecular Biotechnology, Genome Analysis, \*

Beutenbergstrasse 11, D-07745 Jena, Germany,

\* e.mail: gscj-submit@genome.imb-jena.de

\* URL: http://genome.imb-jena.de/

and

\* Keio University School of Medicine, Dept. of Molecular Biology, \*

Tokyo 160-8582, Japan,

\* e.mail: shimizu@db-med.keio.ac.jp

\* URL: http://adenine.dmb.med.keio.ac.jp/

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e.mail:

info.genome@gbf.de

\* URL: http://genome.gbf.de/

and

\* Max-Planck Institute for Molecular Genetics,

\* Innestrasse 73, D-14195 Berlin, Germany,

\* e.mail: info-chr21@molgen.mpg.de

\* URL: http://chr21.rz-berlin.mpg.de/.

Location/Qualifiers

1. 340000

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

<1.10570

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone lib="RPC11,3-5 PAC library"

/note="Accession No. AF129075"

1590. 121668

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="P100J12"

/clone lib="RPC11,3-5 PAC library"

/note="Accession No. AF129075"

source

49591. 190996  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="P79E4"  
/clone lib="RPC11,3-5 PAC library"  
/note="Accession No. AF129075"

source

169378. 274529  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="P84N21"  
/clone lib="RPC11,3-5 PAC library"  
/note="Accession No. AF124730"  
258688. 340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="P292A20, 3' partial"  
/clone lib="RPC11,3-5 PAC library"  
/note="Accession No. AF124731"

source

290605. 340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="P292A20, 3' partial"  
/clone lib="RPC11,3-5 PAC library"  
/note="Accession No. AF124731"

source

340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"  
/clone="P71073, 3' partial"  
/clone lib="CMP21 P1 library"  
/note="Accession No. AP000241"  
complement (1930. 10199)  
/gene="ZNF294"  
complement (join (1930. 2178,3395. 3493,3987. 4190, 10020. 10199))

gene

10020. 10199  
/gene="ZNF294"  
complement (join (1930. 2178,3395. 3493,3987. 4190, 10020. 10199))

CDS

10020. 10199  
/gene="ZNF294"  
/note="mRNA consists of Accession Nos. AB018257 and AF150217, junction confirmed by RT-PCR"  
/codon\_start=1  
/product="containing human KIAA0714 protein"  
/protein\_id="CAB90430.1"  
/db\_xref="GI:7717308"  
/db\_xref="Genbank:13082"  
/db\_xref="GOA:O94822"

10020. 10199  
/db\_xref="UniProt/Swiss-Prot:O94822"  
/translation="MDRVGGKPGSGYRVSSSAARRKLCPCGVNRGLSAQSTAAT  
IDSLVDSFRVLRKLSKQVTKLKAQEFQETGCTETVKGVLVWPRIFCKIS  
LDHRRVREATQQAPEKLUKVKQLAPFLKSLMGYWLMAQCDDTTPPAFAKDAFEA  
APFSPQPEAIAFCDEITSVSSGII"  
complement (1930. 2178)  
/gene="ZNF294"  
/number=4

exon

2434. 2737  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (2915. 3216)  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (3395. 3493)  
/gene="ZNF294"  
/number=3  
complement (3987. 4190)

repeat\_region

2434. 2737  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (2915. 3216)  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (3395. 3493)  
/gene="ZNF294"  
/number=3  
complement (3987. 4190)

exon

2434. 2737  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (2915. 3216)  
/note="AluX"  
/rpt\_family="SINE/Alu"  
/rpt\_type=DISPERSED  
complement (3395. 3493)  
/gene="ZNF294"  
/number=3  
complement (3987. 4190)



\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 702: contig of 702 bp in length  
 \* 703 782: gap of unknown length  
 \* 783 1322: contig of 540 bp in length  
 \* 1323 1402: gap of unknown length  
 \* 1403 84895: contig of 83493 bp in length  
 \* 84896 84975: gap of unknown length  
 \* 84976 85473: contig of 498 bp in length  
 \* 85474 85553: gap of unknown length  
 \* 85554 86109: contig of 556 bp in length  
 \* 86110 86189: gap of unknown length  
 \* 86190 86795: contig of 606 bp in length  
 \* 86796 86875: gap of unknown length  
 \* 86876 87456: contig of 581 bp in length  
 \* 87457 87536: gap of unknown length  
 \* 87537 88144: contig of 608 bp in length  
 \* 88145 88224: gap of unknown length  
 \* 88225 88871: contig of 647 bp in length  
 \* 88872 88951: gap of unknown length  
 \* 88952 89406: contig of 455 bp in length  
 \* 89407 89486: gap of unknown length  
 \* 89487 90083: contig of 597 bp in length  
 \* 90084 90163: gap of unknown length  
 \* 90164 90788: contig of 625 bp in length.

#### FEATURES

source  
 1..90788  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosomes="2"  
 /map="36B3-36B4"  
 /clone="DS07161 (D356)"  
 /clone\_lib="P1 library, partial Sau3A in pAd10sacBII"  
 703..782  
 /estimated\_length=unknown  
 1323..1402  
 /estimated\_length=unknown  
 84896..84975  
 /estimated\_length=unknown  
 84976..85553  
 /estimated\_length=unknown  
 86110..86189  
 /estimated\_length=unknown  
 86796..86875  
 /estimated\_length=unknown  
 87457..87536  
 /estimated\_length=unknown  
 88145..88224  
 /estimated\_length=unknown  
 88872..88951  
 /estimated\_length=unknown  
 89407..89486  
 /estimated\_length=unknown  
 90084..90163  
 /estimated\_length=unknown

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 14; Length 90788;

Best Local Similarity 95.2%; Pred. No. 1.5e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGAAGGAGGA 21

Db 73424 AAAAAACACTTGAAGGAGGA 73444

RESULT 29

AC009250/c

LOCUS

DEFINITION Drosophila melanogaster, Chromosome 2L, region 36B-36B, BAC clone

AC009250 170362 bp DNA linear INV 10-MAR-2001

#### ACCESSION VERSION KEYWORDS SOURCE ORGANISM

BACR27M18, complete sequence.  
 AC009250 GI:13270514  
 HTG.

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 170362)

#### REFERENCE

##### AUTHORS

Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,  
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,  
 Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,  
 Dodson, K., Dorsett, V., Doup, L.E., Doyle, C.C., Dresnek, D., Farfan, D.,  
 Ferreria, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,  
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,  
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
 Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
 Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,  
 Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M.,  
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2L, region 36B-36B

Unpublished

2 (bases 1 to 170362)

##### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.

Direct Submission

##### TITLE

##### JOURNAL

##### COMMENT

Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Mar 10, 2001 this sequence version replaced gi:5734689.  
 Sequence submitted by:  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

#### FEATURES

##### source

1..170362  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="36B-36B"  
 /clone="BACR27M18 (D1027)"  
 /clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcORI in  
 pBACe3.6)"

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 2; Length 170362;

Best Local Similarity 95.2%; Pred. No. 1.4e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGAAGGAGGA 21

Db 2196 AAAAAACACTTGAAGGAGGA 2176

RESULT 30

AC090872	AC090872	178481 bp	DNA	linear	INV 14-MAR-2001
LOCUS	Drosophila melanogaster, chromosome 2L, region 36B-36C, BAC clone				
DEFINITION	BACR06G19, complete sequence.				
ACCESSION	AC090872				
VERSION	AC090872.1	GI:13324747			
KEYWORDS	HTG.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 178481) Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.				
TITLE	Sequencing of Drosophila chromosome 2L, region 36B-36C				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 178481) Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (14-MAR-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US				
COMMENT	Sequence submitted by: Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720				
FEATURES	This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site ( <a href="http://www.fruitfly.org/sequence/">http://www.fruitfly.org/sequence/</a> ) or send email to <a href="mailto:bdgp@fruitfly.berkeley.edu">bdgp@fruitfly.berkeley.edu</a> . Location/Qualifiers 1. 178481 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /strain="y; cn bw sp" /db_xref="taxon:7227" /chromosome="2L" /map="36B-36C" /clone="BACR06G19 (D1335)" /clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"				
ORIGIN	Query Match 77.6%; Score 19.4; DB 2; Length 178481; Best Local Similarity 95.2%; Pred. No. 1.4e+02; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				

QY	1	AAAAAACACTTGGAGGGAGA	21		
DB	27034	AAAAAACACTTGGAGGGAGA	27054		
RESULT 31					
AC152322/c					
LOCUS	AC152322	202907 bp	DNA	linear	HTG 01-JUL-2005
DEFINITION	Bos taurus clone CH240-207, WORKING DRAFT SEQUENCE, 13 unordered pieces.				
ACCESSION	AC152322				
VERSION	AC152322.3	GI:68227244			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Bos taurus (cow)				
ORGANISM	Bos taurus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
AUTHORS	1 (bases 1 to 202907) Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Crease,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falla,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenzshewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwankwelen,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poldexter,A., Popovic,D., Primus,E., Pu,L., L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scheier,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajz,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlsczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 202907)				
AUTHORS	Worley,K.C.				



COMMENT	On Sep 16, 2002 this sequence version replaced gi:10728835.
FEATURES	Location/Qualifiers
source	1..247634
	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/chromosome="2L"
gene	complement (<260..934)
	/gene="CG32832"
	/locus_tag="Dmel_CG32832"
	/old_locus_tag="CG32832"
	/map="36B3-36B3"
	/db_xref="FLYBASE:FBgn0052832"
mRNA	complement (join (<260..535,676..716,770..934))
	/gene="CG32832"
	/locus_tag="Dmel_CG32832"
	/old_locus_tag="CG32832"
	/product="CG32832-PA"
	/db_xref="FLYBASE:FBgn0052832"
CDS	complement (join (260..535,676..716,770..875))
	/gene="CG32832"
	/locus_tag="Dmel_CG32832"
	/old_locus_tag="CG32832"
	/note="CG32832 gene product from transcript CG32832-RA"
	/codon_start=1
	/product="CG32832-PA"
	/protein_id="AA110977.1"
	/db_xref="GI:22946686"
	/db_xref="FLYBASE:FBgn0052832"
	/translation="MSKGTGPLSKLYNITISTIDKFVPGAVQPLWSPAGPRTVTFVFA FAFKWSLVLGLSDTLRSPPANISLNCQCSLAVTGLIWSRYSVITPKYNLLAVNIA VFLIQGLVMVKHLWRSENRNAVFNHSHVPIKSGDW" 1176..7982
gene	/gene="CG31743"
	/locus_tag="Dmel_CG31743"
	/old_locus_tag="CG31743"
	/map="36B3-36B3"
	/db_xref="FLYBASE:FBgn0032618"
mRNA	join (1176..1579,2248..2319,2694..3460,6923..7394,7455..7982)
	/gene="CG31743"
	/locus_tag="Dmel_CG31743"
	/old_locus_tag="CG31743"
	/product="CG31743-RA"
	/db_xref="FLYBASE:FBgn0032618"
CDS	join (3173..3460,6923..7394,7455..7684)
	/gene="CG31743"
	/locus_tag="Dmel_CG31743"
	/old_locus_tag="CG31743"
	/note="CG31743 gene product from transcript CG31743-RA"
	/codon_start=1
	/product="CG31743-PA"
	/protein_id="AAF53584.2"
	/db_xref="GI:22946687"
	/db_xref="FLYBASE:FBgn0032618"
	/translation="MDQLVLSFCISIGFSLGNEFSTGLPFEDGNNNNVKITRRSEL TOTINQRQEFMRQCELDGHTQTLDELSELQMDHIMVDRHEKLLCYVPKACTNM KRYLMLTNKNGTDPLOIPGSLHVSVMFTKLYDLSEAEQQQLSDEYTRFLVRH FLERLLSAYRNKLEQDPSRQVQSRVGRQIVKELRPGASNNLSRGGDVSFGFIQY LVTPELSRANQSDYNEHWEVIAKLCNPKVKNVGVKDYTLDDLSALALYLAGADNLT FPTGHPSSSTRANRYFDPPLPIGAIRKLYDIYEDDDFLRFDYALDEVLGPEFG" complement (8781..9600)
gene	/gene="CG6115"
	/locus_tag="Dmel_CG6115"
	/old_locus_tag="CG6115"
	/map="36B3-36B3"
	/db_xref="FLYBASE:FBgn0040985"
mRNA	complement (join (8781..9336,9452..9600))
	/gene="CG6115"
	/locus_tag="Dmel_CG6115"
	/old_locus_tag="CG6115"
	/product="CG6115-PA"
	/db_xref="FLYBASE:FBgn0040985"

```
CDS
complement(join(9121..9336,9452..9493))
/gene="CG6115"
/locus_tag="Dmel_CG6115"
/old_locus_tag="CG6115"
/note="CG6115 gene product from transcript CG6115-RA"
/codon_start=1
/product="CG6115-PA"
/protein_id="AAF53585.1"
/db_xref="GI:7298358"
/db_xref="FLYBASE:FBgn040985"
/translation="MSQRSKVISLYKHLYLGREYPLGNGPKQFKRQIHDAFMNHKD
EQDPKIVALLAQGRYLAKVEALYSLKRYSKQYSYND"
<10662..>19731
/gene="CG15133"
/locus_tag="Dmel_CG15133"
/old_locus_tag="CG15133"
/map="36B3-36B4"
/db_xref="FLYBASE:FBgn032619"
/join(<10662..10784,11842..12681,13117..13396,13452..16602,
16660..16863,16946..17489,17550..17830,17896..18016,
19147..>19731)
/gene="CG15133"
/locus_tag="Dmel_CG15133"
/old_locus_tag="CG15133"
/product="CG15133-RA"

Query Match      77.6%; Score 19.4; DB 2; Length 247634;
Best Local Similarity 95.2%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGA 21
Db 87164 AAAAAACACTTGGATGGAGA 87184
|||||

RESULT 33
MTMAM1
LOCUS      428 bp mRNA linear ROD 18-APR-2005
DEFINITION Mouse mRNA for gamma-7S nerve growth factor (gamma-NGF) fragment.
ACCESSION X00472.1 GI:54260
VERSION    X00472.1
KEYWORDS   complementary DNA; nerve growth factor; serine protease.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 428)
AUTHORS    Howles,P.N., Dickinson,D.P., DiCaprio,L.L., Woodworth-Gutai,M. and
Gross,K.W.
TITLE      Use of a cDNA recombinant for the gamma-subunit of mouse nerve
growth factor to localize members of this multigene family near the
TAM-1 locus on chromosome 7
JOURNAL    Nucleic Acids Res. 12 (6), 2791-2805 (1984)
PUBMED    6200835
FEATURES   Location/Qualifiers
            source
            1..428
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            <1..408
            /note="unnamed protein product; gamma-NGF"
            /codon_start=1
            /protein_id="CAA25154.1"
            /db_xref="GI:54261"
            /db_xref="GOA:Q62284"
            /db_xref="InterPro:IPR001254"
            /db_xref="UniProt/TREMBL:Q62284"
            /translation="SKPADITDTVKPILTPTEPKLGSTCLASGWSITPTKPFQTTDD
LYCVNLKLPNEDCAKAHIEKVTDMLCAGEMEGKDTCKGDSGGPLICDGVLQGIT5
WGHTPCGEPDMPGVYTKLNKFTSWIKDTWAKNP"

CDS
complement(join(9121..9336,9452..9493))
/gene="CG6115"
/locus_tag="Dmel_CG6115"
/old_locus_tag="CG6115"
/note="CG6115 gene product from transcript CG6115-RA"
/codon_start=1
/product="CG6115-PA"
/protein_id="AAF53585.1"
/db_xref="GI:7298358"
/db_xref="FLYBASE:FBgn040985"
/translation="MSQRSKVISLYKHLYLGREYPLGNGPKQFKRQIHDAFMNHKD
EQDPKIVALLAQGRYLAKVEALYSLKRYSKQYSYND"
<10662..>19731
/gene="CG15133"
/locus_tag="Dmel_CG15133"
/old_locus_tag="CG15133"
/map="36B3-36B4"
/db_xref="FLYBASE:FBgn032619"
/join(<10662..10784,11842..12681,13117..13396,13452..16602,
16660..16863,16946..17489,17550..17830,17896..18016,
19147..>19731)
/gene="CG15133"
/locus_tag="Dmel_CG15133"
/old_locus_tag="CG15133"
/product="CG15133-RA"

Query Match      77.6%; Score 19.4; DB 2; Length 247634;
Best Local Similarity 95.2%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGA 21
Db 87164 AAAAAACACTTGGATGGAGA 87184
|||||

RESULT 33
MTMAM1
LOCUS      428 bp mRNA linear ROD 18-APR-2005
DEFINITION Mouse mRNA for gamma-7S nerve growth factor (gamma-NGF) fragment.
ACCESSION X00472.1 GI:54260
VERSION    X00472.1
KEYWORDS   complementary DNA; nerve growth factor; serine protease.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 428)
AUTHORS    Howles,P.N., Dickinson,D.P., DiCaprio,L.L., Woodworth-Gutai,M. and
Gross,K.W.
TITLE      Use of a cDNA recombinant for the gamma-subunit of mouse nerve
growth factor to localize members of this multigene family near the
TAM-1 locus on chromosome 7
JOURNAL    Nucleic Acids Res. 12 (6), 2791-2805 (1984)
PUBMED    6200835
FEATURES   Location/Qualifiers
            source
            1..428
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            <1..408
            /note="unnamed protein product; gamma-NGF"
            /codon_start=1
            /protein_id="CAA25154.1"
            /db_xref="GI:54261"
            /db_xref="GOA:Q62284"
            /db_xref="InterPro:IPR001254"
            /db_xref="UniProt/TREMBL:Q62284"
            /translation="SKPADITDTVKPILTPTEPKLGSTCLASGWSITPTKPFQTTDD
LYCVNLKLPNEDCAKAHIEKVTDMLCAGEMEGKDTCKGDSGGPLICDGVLQGIT5
WGHTPCGEPDMPGVYTKLNKFTSWIKDTWAKNP"

ORIGIN
```

```
Query Match      76.8%; Score 19.2; DB 9; Length 428;
Best Local Similarity 87.5%; Pred. No. 4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGAAGGGAGACTCA 25
Db 238 AAGAGACACTTGAAGGGTGACTCA 261
|||||

RESULT 34
BVI61004
LOCUS      627 bp DNA linear STS 15-MAY-2004
DEFINITION RPAMMSQ0038539 Roche Palo Alto Mus musculus STS genomic, sequence
tagged site.
ACCESSION BVI61004.1 GI:47264404
VERSION    BVI61004.1
KEYWORDS   STS.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 627)
AUTHORS    Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
TITLE      Mus musculus SNPs
JOURNAL    Unpublished (2003)
COMMENT    Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.
            Location/Qualifiers
            source
            1..627
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /map="1-21087-20591-CAAA01034157.1.1.28512"
            /clone_lib="Roche Palo Alto"
            /note="SNPs developed from assay sequences derived from 15
            different strains of mice (as of October 1, 2003). Those
            strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/OsNj,
            BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
            MRL/MpJ, NZB/BinJ, NZW/BinJ, NZW/Lac, SPRET/Ei."
            <1..>627

STX
ORIGIN
Query Match      76.8%; Score 19.2; DB 10; Length 627;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGACTCA 25
Db 306 ANACAACACTCTGGAAGGGAGACTCA 282
|||||

RESULT 35
RNTXALRC
LOCUS      689 bp mRNA linear ROD 18-APR-2005
DEFINITION Rat tissue mRNA for kallikrein C-terminal region (EC 3.4.21.8).
ACCESSION X03560
VERSION    X03560.1 GI:57370
KEYWORDS   kallikrein; protease; serine protease.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
```



```

REFERENCE
AUTHORS      1 (bases 189 to 689)
TITLE        Gerald, W.L., Chao, J. and Chao, L.
              Immunological identification of rat tissue kallikrein cDNA and
              characterization of the kallikrein gene family
JOURNAL      Biochim. Biophys. Acta 866 (1), 1-14 (1986)
PUBMED       3004582
REFERENCE     2 (bases 1 to 188)
AUTHORS      Chao, L.
TITLE        Direct Submission
JOURNAL      Submitted (13-AUG-1986) Medical University of South Carolina, 171
              Ashley Avenue, Charleston, South Carolina 29425-2211
FEATURES
source       1..689
              Location/Qualifiers
              /organism="Rattus norvegicus"
              /mol_type="mRNA"
              /db_xref="taxon:10118"
              <1..646
              /codon_start=2
              /product="kallikrein"
              /protein_id="CAA27247.1"
              /db_xref="GI:818030"
              /db_xref="GOA:P00758"
              /db_xref="UniProt/Swiss-Prot:P00758"
              /translation="YLCGGVLIDPSWVITAAHCATDNYQVWLGRNNLYDEPPAQHRL
              VSQSFPHGPNQDLIWNHTRPQGDYSNDMLHLGSPADITDGVKVIDLPIEPPKVG
              STCLASGWSITPDGLELSDLOCCVNIIDLLSNEKCVKAHKEVTDMLCAGENDMGKD
              TCKGSGGPLICNGVLGITSWGFNCPGEPKPGIYTKLIKFTPIKVEYKKNP"
polyA_signal 670..675
              /note="putative"
polyA_site   689
ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 689;
Best Local Similarity 87.5%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
Db      476 AAAGACACTTGCAGGGTGACTCA 499

RESULT 36
S48142 LOCUS      777 bp mRNA linear ROD 08-MAY-1993
DEFINITION T-kininogenase=kallikrein homolog [rats, submandibular gland, mRNA,
              777 nt].
ACCESSION   S48142
VERSION     S48142.1 GI:259430
KEYWORDS    Rattus sp.
SOURCE      Rattus sp.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 777)
AUTHORS     Ma, J.X., Chao, J. and Chao, L.
TITLE       Molecular cloning and characterization of rKlk10, a cDNA encoding
              T-kininogenase from rat submandibular gland and kidney
JOURNAL     Biochemistry 31 (44), 10922-10928 (1992)
PUBMED     1420203
REMARK     GenBank staff at the National Library of Medicine created this
              entry [NCBI gisbseq 118095] from the original journal article.
FEATURES
source       1..777
              Location/Qualifiers
              /organism="Rattus sp."
              /mol_type="mRNA"
              /db_xref="taxon:10118"
              1..777
              /gene="T-kininogenase"
              3..737
              /gene="T-kininogenase"
              /note="kallikrein homolog"
              /codon_start=1

```

```

/product="T-kininogenase"
/protein_id="AAB24071.1"
/db_xref="GI:259431"
/translation="DAAPPQGSRIVGGYKCKNSQSPQVQVAINIYELCGVLIDPSKVI
TAACHVSNYHVLLGRNNLFEDEPFAQYRFVNSFPHPDYKFLRNHTRQRGDDYSN
DMLHLSEPADITDGVKVIDLPTPEPKVGSTCLASGWSITKPLANWELPDDLQCVNIH
LLSNEKCIKAEVQKVTDLMLCAGENDRGKDTCKGDSGGPLICDGVLGITSGNVPCA
EPYNGVYTKLIKFTSWIKVEYKKNP"

ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 777;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
Db      567 AAAGACACTTGCAGGGTGACTCA 590

RESULT 37
MMNGFB LOCUS      849 bp mRNA linear ROD 30-MAR-1995
DEFINITION Mouse mRNA for gamma subunit of nerve growth factor (NGF).
ACCESSION   X01389
VERSION     X01389.1 GI:53373
KEYWORDS    Mus musculus (house mouse)
SOURCE      Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 849)
AUTHORS     Ullrich, A., Gray, A., Wood, W.I., Hayflick, J. and Seeburg, P.H.
TITLE       Isolation of a cDNA clone coding for the gamma-subunit of mouse
              nerve growth factor using a high-stringency selection procedure
JOURNAL     DNA 3 (5), 387-392 (1984)
PUBMED     6548955
FEATURES
source       1..849
              Location/Qualifiers
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              18..803
              /note="unnamed protein product; gamma NGF precursor"
              /codon_start=1
              /protein_id="CAA25645.1"
              /db_xref="GI:53374"
              /db_xref="GOA:P00756"
              /translation="MMFLILFLALSLGGIDAAPVQSVQSRIVGGFKCKNSQPMWVAVYR
              YTCYLCGGVLIDPNWVLTAAHCYDDNYKVLGKNLFDKSPSAQHRFVSKAIPHPGFN
              MSIMRKHIRFLEYDYSNDMLLRLSKPADITDVKPTLPTPEPKLGSTCLASGWSI
              TPTKFOPTDGLYCVNLKLLPNEDCAKAIKVTDMLCAGEMDGGKDTCKGDSGGPLI
              CDVLGQITSGWHTPCGEPMPGVYTKLNKFTSWIKDTMAKNP"
              18..89
              /notes="put. signal peptide"
              90..800
              /product="gamma NGF"
              90..388
              /notes="proteolytic peptide B1"
              389..400
              /notes="connecting peptide"
              401..800
              /note="proteolytic peptide C2"
              825..830
              /notes="polyA signal"
              849
              /note="polyA site"

ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 849;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
DB      633 AAAGACACTTGCAGGGTGACTCA 656

RESULT 38
RATPKLK
DEFINITION
  rat pancreatic preprokallikrein mrna.
ACCESSION
  J00758
VERSION
  J00758.1 GI:206200
KEYWORDS
  kallikrein; protease; serine protease.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
  1 (bases 1 to 867)
AUTHORS
  Swift,G.H., Dgorn,J.C., Ashley,P.L., Cummings,S.W. and
  MacDonald,R.J.
TITLE
  Rat pancreatic kallikrein mRNA: nucleotide sequence and amino acid
  sequence of the encoded preproenzyme
JOURNAL
  Proc. Natl. Acad. Sci. U.S.A. 79 (23), 7263-7267 (1982)
PUBMED
  6961406
COMMENT
  Original source text: rat pancreatic tissue.
FEATURES
  source
    1..867
      /organism="Rattus norvegicus"
      /mol_type="mRNA"
      /db_xref="taxon:10116"
ORIGIN

Query Match      76.8%; Score 19.2; DB 9; Length 867;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
DB      658 AAAGACACTTGCAGGGTGACTCA 681

RESULT 40
BC078784
LOCUS
DEFINITION
  Rattus norvegicus nerve growth factor, gamma, mRNA (cDNA clone
  MGC:93329 IMAGE:7128287), complete cds.
ACCESSION
  BC078784
VERSION
  BC078784.1 GI:51261179
KEYWORDS
  MGC.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
  1 (bases 1 to 879)
AUTHORS
  Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
  Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg
  B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
  Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
  GM, Hong L, Stapleton M, Soares MB, Donald MP, Casavant TL,
  Scheetz TE, Brownstein MJ, Udwin TB, Toshiyuki S, Carninci P,
  Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mulahy
  SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
  Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
  Villalón DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
  Helton E, Kettelman M, Madan A, Rodriguez S, Sanchez A, Whitting M,
  Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
  Touchman JW, Green ED, Dickinson MC, Rodriguez AC, Grimwood J,
  Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
  Smalusz DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 879)
DIRECTOR MGC Project.
DIRECT SUBMISSION
DIRECT SUBMISSION
JOURNAL
  Submitted (02-AUG-2004) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
REFERENCE
  NIH-MGC Project URL: http://mgc.nci.nih.gov
CONTACT: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
  R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 184 Row: C Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 13994130.

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
DB      633 AAAGACACTTGCAGGGTGACTCA 656

RESULT 38
RATPKLK
DEFINITION
  rat pancreatic preprokallikrein mrna.
ACCESSION
  J00758
VERSION
  J00758.1 GI:206200
KEYWORDS
  kallikrein; protease; serine protease.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
  1 (bases 1 to 867)
AUTHORS
  Swift,G.H., Dgorn,J.C., Ashley,P.L., Cummings,S.W. and
  MacDonald,R.J.
TITLE
  Rat pancreatic kallikrein mRNA: nucleotide sequence and amino acid
  sequence of the encoded preproenzyme
JOURNAL
  Proc. Natl. Acad. Sci. U.S.A. 79 (23), 7263-7267 (1982)
PUBMED
  6961406
COMMENT
  Original source text: rat pancreatic tissue.
FEATURES
  source
    1..867
      /organism="Rattus norvegicus"
      /mol_type="mRNA"
      /db_xref="taxon:10116"
ORIGIN

Query Match      76.8%; Score 19.2; DB 9; Length 867;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTTGGAGGGAGACTCA 25
      ||| ||||| ||||| ||||| |||||
DB      652 AAAGACACTTGCAGGGTGACTCA 675

RESULT 39
RATKALPS
LOCUS
DEFINITION
  Rat submaxillary gland PS kallikrein mRNA, complete cds.
ACCESSION
  M11563
VERSION
  M11563.1 GI:205029
KEYWORDS
  Rattus norvegicus (Norway rat)
SOURCE
  Rattus norvegicus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
  1 (bases 1 to 873)
AUTHORS
  Ashley,P.L. and MacDonald,R.J.
TITLE
  Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide
  sequences of four distinct types including tonin
JOURNAL
  Biochemistry 24 (17), 4512-4520 (1985)
PUBMED
  2998455
COMMENT
  Original source text: Rat submaxillary gland, cDNA to mRNA.
FEATURES
  source
    1..873
      /organism="Rattus norvegicus"
      /mol_type="mRNA"
      /db_xref="taxon:10116"
    31..828
      /note="PS kallikrein"
      /codon_start=1
      /protein_id="AAA41464.1"
      /db_xref="GI:205030"
      /translation="MPVTWFLILFLALSLGRNDAPPVQSRVVGNGNCMNSQPQVQV
      AVTFEYLCGGVLIDPSVITAAHCATQNYQVWLGRNNLYDEBPQAHLVSQSPFH
      CDS

```

FEATURES	Location/Qualifiers	CDS	ORIGIN
source	1..879		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	/clone="MGC:93329 IMAGE:7128287"		
	/tissue_type="Kidney, rat (Brown Norway)"		
	/clone_lib="NIH MGC_236"		
	/lab_host="DH10B"		
	/notes="Vector: pExpress1"		
gene	1..879		
	/gene="Ngfg"		
	/notes="synonyms: KUK1, Kkl1c1, rGK-1, preprokallikrein"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
CDS	13..810		
	/gene="Ngfg"		
	/codon_start=1		
	/product="nerve growth factor, gamma"		
	/protein_id="AAH78784.1"		
	/db_xref="GI:51261180"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
	/translations="MPVTWMLFLFLALSIGRNDAAAPPVQSRVVGYNCEMNSQPWQV AVYFGEYLCGVLIDPSSWITIAAHCATDNYQVWLGNNLYDEPPFAQRLVQSFPFH PGFNQDLIWNHTRPQGDVSDNMLLHLSOPADITGVKVIDLPIEPKRVGSTCLASG WGSTPDGLSDSLQCVNIDLLSNKCVKAEHKEEVTDLMLCAGEMDGKDKCKGDSG GPLICNGVLGITSWGNPCGPEKPKGIYTKLIKIFTPWIKVYMKNP"		
ORIGIN			
Query Match	76.8%; Score 19.2; DB 9; Length 879;		
Best Local Similarity	87.5%; Pred. No. 3.6e+02;		
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Qy	2 AAAAAACACTTGGGAAGGAGACTCA 25 		
Db	640 AAAGACACTTGCAGGAGTCA 663 		
RESULT 41			
PHRCL52/c			
LOCUS	1420 bp DNA linear PLN 18-APR-2005		
DEFINITION	P.hysopifolia chloroplast rbcL gene for rubisco (large subunit) (partial).		
ACCESSION	Z37432		
VERSION	Z37432.1 GI:671608		
KEYWORDS	rbcL gene; Rubisco.		
SOURCE	chloroplast Prunella hyssopifolia		
ORGANISM	Prunella hyssopifolia		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Lamiales; Lamiaceae; Nepetoideae; Nepeteae; Prunella.		
REFERENCE	1 (bases 1 to 1420)		
AUTHORS	Kaufmann,M. and Wink,M.		
TITLE	Molecular systematics of the Nepetoideae (family Labiatae): phylogenetic implications from rbcL gene sequences		
JOURNAL	Biosci. Rep. 49, 635-645 (1994)		
REFERENCE	2 (bases 1 to 1420)		
AUTHORS	Kaufmann,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-SEP-1994) M.Kaufmann, Institut fuer Pharmazeutische Biologie, Im Neuenheimer Feld 364, 69121 Heidelberg, Germany		
FEATURES	Location/Qualifiers		
source	1..1420		
	/organism="Prunella hyssopifolia"		
	/organella="plastid:chloroplast"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:39357"		
gene	1..>1420		
	/gene="rbcL"		

FEATURES	Location/Qualifiers	CDS	ORIGIN
source	1..879		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	/clone="MGC:93329 IMAGE:7128287"		
	/tissue_type="Kidney, rat (Brown Norway)"		
	/clone_lib="NIH MGC_236"		
	/lab_host="DH10B"		
	/notes="Vector: pExpress1"		
gene	1..879		
	/gene="Ngfg"		
	/notes="synonyms: KUK1, Kkl1c1, rGK-1, preprokallikrein"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
CDS	13..810		
	/gene="Ngfg"		
	/codon_start=1		
	/product="nerve growth factor, gamma"		
	/protein_id="AAH78784.1"		
	/db_xref="GI:51261180"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
	/translations="MPVTWMLFLFLALSIGRNDAAAPPVQSRVVGYNCEMNSQPWQV AVYFGEYLCGVLIDPSSWITIAAHCATDNYQVWLGNNLYDEPPFAQRLVQSFPFH PGFNQDLIWNHTRPQGDVSDNMLLHLSOPADITGVKVIDLPIEPKRVGSTCLASG WGSTPDGLSDSLQCVNIDLLSNKCVKAEHKEEVTDLMLCAGEMDGKDKCKGDSG GPLICNGVLGITSWGNPCGPEKPKGIYTKLIKIFTPWIKVYMKNP"		
ORIGIN			
Query Match	76.8%; Score 19.2; DB 9; Length 879;		
Best Local Similarity	87.5%; Pred. No. 3.6e+02;		
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Qy	2 AAAAAACACTTGGGAAGGAGACTCA 25 		
Db	640 AAAGACACTTGCAGGAGTCA 663 		
RESULT 41			
PHRCL52/c			
LOCUS	1420 bp DNA linear PLN 18-APR-2005		
DEFINITION	P.hysopifolia chloroplast rbcL gene for rubisco (large subunit) (partial).		
ACCESSION	Z37432		
VERSION	Z37432.1 GI:671608		
KEYWORDS	rbcL gene; Rubisco.		
SOURCE	chloroplast Prunella hyssopifolia		
ORGANISM	Prunella hyssopifolia		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Lamiales; Lamiaceae; Nepetoideae; Nepeteae; Prunella.		
REFERENCE	1 (bases 1 to 1420)		
AUTHORS	Kaufmann,M. and Wink,M.		
TITLE	Molecular systematics of the Nepetoideae (family Labiatae): phylogenetic implications from rbcL gene sequences		
JOURNAL	Biosci. Rep. 49, 635-645 (1994)		
REFERENCE	2 (bases 1 to 1420)		
AUTHORS	Kaufmann,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-SEP-1994) M.Kaufmann, Institut fuer Pharmazeutische Biologie, Im Neuenheimer Feld 364, 69121 Heidelberg, Germany		
FEATURES	Location/Qualifiers		
source	1..1420		
	/organism="Prunella hyssopifolia"		
	/organella="plastid:chloroplast"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:39357"		
gene	1..>1420		
	/gene="rbcL"		

FEATURES	Location/Qualifiers	CDS	ORIGIN
source	1..879		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	/clone="MGC:93329 IMAGE:7128287"		
	/tissue_type="Kidney, rat (Brown Norway)"		
	/clone_lib="NIH MGC_236"		
	/lab_host="DH10B"		
	/notes="Vector: pExpress1"		
gene	1..879		
	/gene="Ngfg"		
	/notes="synonyms: KUK1, Kkl1c1, rGK-1, preprokallikrein"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
CDS	13..810		
	/gene="Ngfg"		
	/codon_start=1		
	/product="nerve growth factor, gamma"		
	/protein_id="AAH78784.1"		
	/db_xref="GI:51261180"		
	/db_xref="GeneID:24594"		
	/db_xref="RATMAP:34215"		
	/db_xref="RGD:3175"		
	/translations="MPVTWMLFLFLALSIGRNDAAAPPVQSRVVGYNCEMNSQPWQV AVYFGEYLCGVLIDPSSWITIAAHCATDNYQVWLGNNLYDEPPFAQRLVQSFPFH PGFNQDLIWNHTRPQGDVSDNMLLHLSOPADITGVKVIDLPIEPKRVGSTCLASG WGSTPDGLSDSLQCVNIDLLSNKCVKAEHKEEVTDLMLCAGEMDGKDKCKGDSG GPLICNGVLGITSWGNPCGPEKPKGIYTKLIKIFTPWIKVYMKNP"		
ORIGIN			
Query Match	76.8%; Score 19.2; DB 15; Length 1420;		
Best Local Similarity	87.5%; Pred. No. 3.4e+02;		
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Qy	2 AAAAAACACTTGGGAAGGAGACTCA 25 		
Db	1125 AATAACACTTGCAGGAGACCCA 1102 		
RESULT 42			
AB036705			
LOCUS	46235 bp DNA linear PRI 06-JAN-2004		
DEFINITION	Homo sapiens EMS1 gene for cortactin, partial cds.		
ACCESSION	AB036705		
VERSION	AB036705.1 GI:40645043		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	1 Yamashita,A. and Togashi,S. Human EMS1(cortactin) gene TITLE Published Only in Database (2004)		
AUTHORS	Yamashita,A. and Togashi,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-JAN-2000) Shin Togashi, Kitasato University, School of Science, Dept. of Biosciences, Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555, Japan (E-mail:stogashi@jet.sci.kitasato-u.ac.jp, Tel:81-42-778-9481, Fax:81-42-778-9481)		
FEATURES	Location/Qualifiers		
source	1..46235		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
gene	join(15820..15906,16040..16113,18353..18482,23064..23117) 24185..24239,25535..25645,28268..28378,28922..29032, 29993..30103,31463..31518		
	/gene="EMS1"		
	join(15820..15906,16040..16113,18353..18482,23064..23117) 24185..24239,25535..25645,28268..28378,28922..29032, 29993..30103,31463..31518		
CDS	join(15820..15906,16040..16113,18353..18482,23064..23117) 24185..24239,25535..25645,28268..28378,28922..29032, 29993..30103,31463..31518		
	/gene="EMS1"		
	/codon_start=1		
	/product="cortactin"		
	/protein_id="BAD06416.1"		
	/db_xref="GI:40645044"		
	/translation="NMKASAGHAVSIADODAGADDWETDPDFVNDVSEKEQRWGAQTV QSGHQEHNHKLBNVFOHQTLKELEGTGPKASHGCGKGVGDMDKSAVGH EYQSKLSKCSQVDSVRGFGKGVMDRVDQSAVGFEYQGTKEKHAQSOKYSSGFGG		

KYGVQADRVDSAVGQYKQTEKHESORDYSKFGGKVGIDKDKVDKSAVGFYQOGK  
TEKHESQKDYKVGFGKGVQVDRQDKALGWDHDEKQLQKHESQKDYKTFGGKRGVQ  
SERQDSAAVGFDPYKEKLAKHESQQDYSKFGGKVGKQDRMDK\*

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 46235;  
Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 2 AAAAAACACTTGAAGGGAGACTCA 25  
|||||  
Db 42251 AAAAGCACTTGGAGGGAGACCCA 42274  
|||||

## RESULT 43

AC101212 59388 bp DNA linear HTG 23-NOV-2001  
Mus musculus clone RP23-182G21, LOW-PASS SEQUENCE SAMPLING.

## AC101212

## DEFINITION

## AC101212

## AC101212

## VERSION

## AC101212.1

## GI:17059987

## HTG; HTGS PHASE0.

## KEYWORDS

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

## Mus musculus

\* 26466 26565: gap of 100 bp  
\* 26566 27256: contig of 691 bp in length  
\* 27257 27356: gap of 100 bp  
\* 27357 28058: contig of 702 bp in length  
\* 28059 28158: gap of 100 bp  
\* 28159 28837: contig of 679 bp in length  
\* 28838 28937: gap of 100 bp  
\* 28938 29646: contig of 709 bp in length  
\* 29647 29746: gap of 100 bp  
\* 29747 30435: contig of 689 bp in length  
\* 30436 30535: gap of 100 bp  
\* 30536 31197: contig of 662 bp in length  
\* 31198 31297: gap of 100 bp  
\* 31298 31960: contig of 663 bp in length  
\* 31961 32060: gap of 100 bp  
\* 32061 32744: contig of 664 bp in length  
\* 32745 32824: gap of 100 bp  
\* 32825 33496: contig of 672 bp in length  
\* 33497 33596: gap of 100 bp  
\* 33597 34285: contig of 689 bp in length  
\* 34286 34385: gap of 100 bp  
\* 34386 35068: contig of 683 bp in length  
\* 35069 35168: gap of 100 bp  
\* 35169 35847: contig of 679 bp in length  
\* 35848 35947: gap of 100 bp  
\* 35948 36632: contig of 685 bp in length  
\* 36633 36732: gap of 100 bp  
\* 36733 37397: contig of 665 bp in length  
\* 37398 37497: gap of 100 bp  
\* 37498 38193: contig of 696 bp in length  
\* 38194 38293: gap of 100 bp  
\* 38294 39010: contig of 717 bp in length  
\* 39011 39110: gap of 100 bp  
\* 39111 39797: contig of 687 bp in length  
\* 39798 39897: gap of 100 bp  
\* 39898 40560: contig of 663 bp in length  
\* 40561 40660: gap of 100 bp  
\* 40661 41347: contig of 687 bp in length  
\* 41348 41447: gap of 100 bp  
\* 41448 42132: contig of 685 bp in length  
\* 42133 42232: gap of 100 bp  
\* 42233 42917: contig of 685 bp in length  
\* 42918 43017: gap of 100 bp  
\* 43018 43718: contig of 701 bp in length  
\* 43719 43818: gap of 100 bp  
\* 43819 44513: contig of 695 bp in length  
\* 44514 44613: gap of 100 bp  
\* 44614 45293: contig of 680 bp in length  
\* 45294 45393: gap of 100 bp  
\* 45394 46119: contig of 726 bp in length  
\* 46120 46219: gap of 100 bp  
\* 46220 46868: contig of 649 bp in length  
\* 46869 46968: gap of 100 bp  
\* 46969 47625: contig of 657 bp in length  
\* 47626 47725: gap of 100 bp  
\* 47726 48389: contig of 664 bp in length  
\* 48390 48489: gap of 100 bp  
\* 48490 49170: contig of 681 bp in length  
\* 49171 49270: gap of 100 bp  
\* 49271 49968: contig of 698 bp in length  
\* 49969 50068: gap of 100 bp  
\* 50070 50849: contig of 681 bp in length  
\* 50850 51561: contig of 712 bp in length  
\* 51562 52311: contig of 100 bp  
\* 52312 52411: contig of 650 bp in length  
\* 52412 53116: gap of 100 bp  
\* 53117 53216: contig of 705 bp in length  
\* 53217 53934: contig of 718 bp in length

Query Match 76.8%; Score 19.2; DB 14; Length 59388;  
Best Local Similarity 87.5%; Pred. No. 2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACACTTGGAGGGAGACTCA 25  
|||||  
Db 13725 AAAAACACTTGGACAGAGACTGA 13748  
|||||

RESULT 44  
AC123415.2  
WPCOMMENT

Sequence split into 5 fragments LOCUS AC123415 Accession AC123415

Fragment Name	Begin	End
AC123415_0	1	110000
AC123415_1	100001	210000
AC123415_2	200001	310000
AC123415_3	300001	410000
AC123415_4	400001	421216

Continuation (3 of 5) of AC123415 from base 200001 (AC123415 Rattus norvegicus clone CH

Query Match 76.8%; Score 19.2; DB 14; Length 110000;

Best Local Similarity 87.5%; Pred. No. 1.9e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACTCA 25  
|||||  
Db 50734 AAAAACACTTGGAAAGGTGATTCA 50757  
|||||

RESULT 45

AC159696

LOCUS

DEFINITION Bos taurus clone CH240-75G13, WORKING DRAFT SEQUENCE, 9 unordered pieces.

ACCESSION AC159696

VERSION AC159696.3 GI:68299868

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

1 (bases 1 to 125379)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Ceasar, H., Center, A., Cardenas, V., Carter, K., Cavazos, I., Chen, R., Chen, Y., Chen, Z., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., Dederich, D., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dinu, H., Divya, K., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Duval, B., Eaves, K., Egan, A., Escotto, M., Evans, C. A., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamill, C., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, N., Hernandez, J., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jarkpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Jolivet, A., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pflannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H., Sanders, W., Savery, G., Scher, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Slasson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 125379)  
 Worley, K.C.  
 Direct Submission  
 Submitted (07-APR-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 125379)  
 Cow Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (01-JUN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jun 29, 2005 this sequence version replaced gi:62420179.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: FELK  
 Center clone name: CH240-75G13  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 122732 bases at least Q40  
 Consensus quality: 123561 bases at least Q30  
 Consensus quality: 124261 bases at least Q20  
 Estimated insert size: 124753; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 28774: contig of 28774 bp in length  
 \* 28775 28824: gap of 50 bp

Japan (E-mail:hattori@sc.riken.go.jp,  
URL:http://hgp.qsc.riken.go.jp/, Tel:81-42-778-9923,  
Fax:81-42-778-9924)  
On May 31, 2000 this sequence version replaced gi:6997436.

## COMMENT

----- Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hgp.qsc.riken.go.jp/  
Contact: hattori@sc.riken.go.jp  
----- Project Information  
Center project name: Humdraft11  
Center clone name: CMB9-16D10  
----- Summary Statistics

Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 122741 bases at least Q40  
Consensus quality: 131182 bases at least Q30  
Consensus quality: 135698 bases at least Q20  
Insert size: 138675; sum-of-contigs  
Quality coverage: 4.07x in Q20 bases; sum-of-contigs

-----  
NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 24550 contig of 24550 bp in length
24651 34442 contig of 9792 bp in length
34543 43950 contig of 9408 bp in length
44051 51303 contig of 7253 bp in length
51404 58548 contig of 7145 bp in length
58649 64379 contig of 5731 bp in length
64480 71839 contig of 7360 bp in length
71940 76792 contig of 4853 bp in length
76893 82501 contig of 5609 bp in length
82602 86341 contig of 3740 bp in length
86442 91801 contig of 100 bp
91801 97119 contig of 5319 bp in length
97120 97220 contig of 3777 bp in length
100996 100996 contig of 100 bp
100997 101097 contig of 5261 bp in length
101097 106358 contig of 100 bp
106358 110874 contig of 4417 bp in length
110874 110975 contig of 100 bp
110975 114227 contig of 3453 bp in length
114227 114528 contig of 100 bp
114528 118179 contig of 3652 bp in length
118179 118280 contig of 100 bp
118280 121482 contig of 3203 bp in length
121482 121582 contig of 100 bp
121582 124342 contig of 2760 bp in length
124342 124443 contig of 100 bp
124443 126512 contig of 2070 bp in length
126512 126613 contig of 100 bp
126613 129400 contig of 2788 bp in length
129400 129501 contig of 100 bp
129501 131165 contig of 1665 bp in length
131165 131266 contig of 100 bp
131266 132520 contig of 1255 bp in length
132520 132621 contig of 100 bp
132621 133646 contig of 1026 bp in length
133646 133746 contig of 100 bp
133746 134903 contig of 1157 bp in length
134903 135003 contig of 1039 bp in length
135003 136042 contig of 100 bp
136042 136142 contig of 100 bp
136142 137315 contig of 1173 bp in length
137315 137415 contig of 100 bp
137415 138609 contig of 1194 bp in length
138609 138709 contig of 100 bp
138709 139892 contig of 1183 bp in length
139892 139993 contig of 100 bp
139993 141097 contig of 1105 bp in length
141097 141198 contig of 100 bp
141198 141675 contig of 478 bp in length.
```

## FEATURES

## source

```
1. 141675
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="11"
   /map="11q13"
   /clone="CMB9-16D10"
1. 24550
   /note="assembly_fragment"
24651..34442
   /note="assembly_fragment"
34543..43950
   /note="assembly_fragment"
44051..51303
   /note="assembly_fragment"
```

## misc\_feature

```
1..24550
```

## misc\_feature

```
24651..34442
```

## misc\_feature

```
34543..43950
```

## misc\_feature

```
44051..51303
```

Sequence updated (12-Oct-1999)  
Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 24550: contig of 24550 bp in length
24551 24650: gap of 100 bp
24651 34442: contig of 9792 bp in length
```





AC161483  
AC161483.1 GI:65301498  
HTG: HTGS PHASE1; HTGS DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 161497)  
AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
TITLE Mus musculus chromosome 5, clone RP23-469M13  
REFERENCE 2 (bases 1 to 161497)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, M., Anderson, S., Atarachi, H.M., Barna, N., Bastien, V.,  
Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J.,  
Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
DeArelano, K., Diaz, J.S., Dodge, S., FitzGerald, M., Gage, D.,  
Ericksen, J., Faro, S., Ferreira, P., FitzGerald, M., Hafez, N.,  
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
McClean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
Topkan, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,  
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (15-MAY-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Broad Institute of MIT and Harvard  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: L32735  
Center clone name: 469 M 13  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 152487 bases at least Q40  
Consensus quality: 156020 bases at least Q30  
Consensus quality: 158188 bases at least Q20  
Insert size: 230000; agarose-fp  
Insert size: 159497; sum-of-contigs  
Quality coverage: 3.7 in Q20 bases; agarose-fp  
Quality coverage: 5.3 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1388: contig of 1388 bp in length  
\* 1389 1488: gap of unknown length  
\* 1489 1812: contig of 324 bp in length  
\* 1813 1912: gap of unknown length  
\* 1913 2674: contig of 762 bp in length  
\* 2675 2775: gap of unknown length  
\* 2775 3270: contig of 496 bp in length  
\* 3271 3370: gap of unknown length  
\* 3371 4460: contig of 1090 bp in length  
\* 4461 4560: gap of unknown length  
\* 4561 6908: contig of 2348 bp in length  
\* 6909 7008: gap of unknown length  
\* 7009 9982: contig of 2974 bp in length  
\* 9983 10083: gap of unknown length  
\* 10084 11832: contig of 1750 bp in length  
\* 11833 11933: gap of unknown length  
\* 11934 14904: contig of 2971 bp in length  
\* 14905 15004: gap of unknown length  
\* 15005 18537: contig of 3534 bp in length  
\* 18538 18638: gap of unknown length  
\* 18639 21921: contig of 3284 bp in length  
\* 21922 26146: gap of unknown length  
\* 26147 31488: contig of 5243 bp in length  
\* 31489 31588: gap of unknown length  
\* 31589 36567: contig of 4979 bp in length  
\* 36568 42840: contig of 6173 bp in length  
\* 42841 42940: gap of unknown length  
\* 42941 70982: contig of 28042 bp in length  
\* 70983 87513: contig of 16431 bp in length  
\* 87514 104741: contig of 17127 bp in length  
\* 104742 104840: gap of unknown length  
\* 104841 128106: contig of 23266 bp in length  
\* 128107 128206: gap of unknown length  
\* 128207 154032: contig of 25826 bp in length  
\* 154033 154133: gap of unknown length  
\* 154134 161497: contig of 7365 bp in length.  
FEATURES  
source  
1..161497  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="5"  
/map="5"  
/clone\_lib="RP23-469M13"  
/clone="RP23-469M13"  
/notes="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
1389..1488  
/estimated\_length=unknown  
1489..1812  
/notes="assembly\_fragment"  
1813..1912  
/estimated\_length=unknown  
1913..2674  
/notes="assembly\_fragment"  
2675..2774  
/estimated\_length=unknown  
2775..3270  
/notes="assembly\_fragment"  
3271..3370  
/estimated\_length=unknown  
3371..4460  
/notes="assembly\_fragment"  
4461..4560  
/estimated\_length=unknown  
4561..6908  
/notes="assembly\_fragment"  
6909..7008  
/estimated\_length=unknown  
7009..9982  
/notes="assembly\_fragment"  
misc\_feature  
1..1388  
/notes="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
1389..1488  
/estimated\_length=unknown  
1489..1812  
/notes="assembly\_fragment"  
1813..1912  
/estimated\_length=unknown  
1913..2674  
/notes="assembly\_fragment"  
2675..2774  
/estimated\_length=unknown  
2775..3270  
/notes="assembly\_fragment"  
3271..3370  
/estimated\_length=unknown  
3371..4460  
/notes="assembly\_fragment"  
4461..4560  
/estimated\_length=unknown  
4561..6908  
/notes="assembly\_fragment"  
6909..7008  
/estimated\_length=unknown  
7009..9982  
/notes="assembly\_fragment"

gap 9983..10082  
/estimated\_length=unknown  
misc\_feature 10083..11832  
/note="assembly\_fragment"  
gap 11833..11932  
/estimated\_length=unknown  
misc\_feature 11933..14903  
/note="assembly\_fragment"  
gap 14904..15003  
/estimated\_length=unknown  
misc\_feature 15004..18537  
/note="assembly\_fragment"  
gap 18538..18637  
/estimated\_length=unknown  
misc\_feature 18638..21921  
/note="assembly\_fragment"  
gap 21922..22021  
/estimated\_length=unknown  
misc\_feature 22022..26145  
/note="assembly\_fragment"  
gap 26146..26245  
/estimated\_length=unknown  
misc\_feature 26246..31488  
/note="assembly\_fragment"  
gap 31489..31588  
/estimated\_length=unknown  
misc\_feature 31589..33657  
/note="assembly\_fragment"  
gap 33658..36667  
/estimated\_length=unknown  
misc\_feature 36668..42840  
/note="assembly\_fragment"  
gap 42841..42940  
/estimated\_length=unknown  
misc\_feature 42941..70982  
/note="assembly\_fragment"  
gap 70983..71082  
/estimated\_length=unknown  
misc\_feature 71083..87513  
/note="assembly\_fragment"  
gap 87514..87613  
/estimated\_length=unknown  
misc\_feature 87614..104740  
/note="assembly\_fragment"  
gap 104741..104840  
/estimated\_length=unknown  
misc\_feature 104841..128106  
/note="assembly\_fragment"  
gap 128107..128206  
/estimated\_length=unknown  
misc\_feature 128207..154032

Query Match 76.8%; Score 19.2; DB 14; Length 161497;  
Best Local Similarity 87.5%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAACACTTGGAGGAGACTC 24 linear ROD 04-SRP-2004  
||||| 162542 bp DNA  
Db 78454 AAAAAACACTGGGAGTGATCTC 78477  
||||| 162542 bp DNA  
  
RESULT 49  
AC114655/c  
LOCUS  
DEFINITION Mus musculus chromosome 1, clone RP24-160J12, complete sequence.  
ACCESSION AC114655  
VERSION AC114655.12 GI:51898957  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 162542)  
Birren,B., Nusbaum,C. and Lander,E.  
Mus musculus chromosome 1, clone RP24-160J12  
Unpublished  
2 (bases 1 to 162542)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., R.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 162542)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (03-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 162542)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (03-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 162542)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (03-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 162542)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (03-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 162542)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (03-AUG-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
 Zimmer, A. and Zody, M.

## TITLE

Direct Submission  
 Submitted (04-SEP-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 4, 2004 this sequence version replaced gi:50897427.  
 All repeats were identified using RepeatMasker:  
 Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WTHR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@broad.mit.edu

----- Project Information

Center Project name: L24414

Center clone name: 160\_J\_12

## FEATURES

## source

## Location/Qualifiers

1..162542  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="1"  
 /clone="RP24-160J12"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

## misc\_feature

1..4  
 /notes="clone boundary  
 clone end:SP6  
 site:MboI"

## repeat\_region

complement(2701..2910)

## repeat\_region

/rpt\_family="Lx3"

## repeat\_region

complement(2912..3061)

## repeat\_region

/rpt\_family="L1M1"

## repeat\_region

3243..3832

## repeat\_region

/rpt\_family="Lx9"

## repeat\_region

3833..4198

## repeat\_region

/rpt\_family="RMER19B"

## repeat\_region

4404..4472

## repeat\_region

/rpt\_family="T-rich"

## repeat\_region

complement(4486..4610)

## repeat\_region

/rpt\_family="B1F"

## repeat\_region

complement(5657..5811)

## repeat\_region

/rpt\_family="RSINE1"

## repeat\_region

5991..6234

## repeat\_region

/rpt\_family="ORR1B1"

## repeat\_region

complement(6237..6315)

## repeat\_region

/rpt\_family="L1\_MM"

## repeat\_region

6313..6557

## repeat\_region

/rpt\_family="L1\_MM"

## repeat\_region

6637..6692

## repeat\_region

/rpt\_family="ORR1B1"

## repeat\_region

complement(6749..6870)

## repeat\_region

/rpt\_family="B1F"

## repeat\_region

7031..7231

## repeat\_region

/rpt\_family="MT2B"

## repeat\_region

complement(7225..7284)

## repeat\_region

/rpt\_family="ORR1D"

## repeat\_region

7285..7346

## repeat\_region

/rpt\_family="TD\_B1"

## repeat\_region

7298..7356

## repeat\_region

/rpt\_family="PB1D7"

## repeat\_region

complement(7440..7651)

## repeat\_region

/rpt\_family="B3A"

## repeat\_region

8046..8188

## repeat\_region

/rpt\_family="B1\_MM"

## repeat\_region

8191..8259

## repeat\_region

/rpt\_family="(A)n"

## repeat\_region

complement(8751..9118)

/rpt\_family="MTC"

repeat\_region 9927..10057  
 /rpt\_family="B1F"  
 repeat\_region 10092..10187  
 /rpt\_family="B1F"  
 repeat\_region complement(10341..10493)  
 /rpt\_family="B1\_MM"  
 repeat\_region 10714..10938  
 /rpt\_family="B3"  
 repeat\_region 11020..11114  
 /rpt\_family="PB1D7"  
 repeat\_region complement(11115..11428)  
 /rpt\_family="MTB"  
 repeat\_region complement(11780..11988)  
 /rpt\_family="RMER4B"  
 repeat\_region 11989..12018  
 /rpt\_family="(TTA)n"  
 repeat\_region complement(12019..12491)  
 /rpt\_family="Lx2"  
 repeat\_region complement(12493..12628)  
 /rpt\_family="RMER4B"  
 repeat\_region complement(12723..13006)  
 /rpt\_family="RMER4B"  
 repeat\_region complement(13033..13086)  
 /rpt\_family="PB1D9"  
 repeat\_region 13101..13238  
 /rpt\_family="CT-rich"  
 repeat\_region complement(13239..13371)  
 /rpt\_family="B1F"  
 repeat\_region complement(13377..13451)  
 /rpt\_family="PB1D7"  
 repeat\_region complement(13571..15097)  
 /rpt\_family="L1\_MM"  
 repeat\_region complement(15213..15717)  
 /rpt\_family="L1"  
 repeat\_region complement(15727..18178)  
 /rpt\_family="LLME"  
 repeat\_region 18179..18213  
 /rpt\_family="(TG)n"  
 repeat\_region complement(18214..18820)  
 /rpt\_family="LLME"  
 repeat\_region 20168..20317  
 /rpt\_family="B1F"  
 repeat\_region complement(20523..20672)  
 /rpt\_family="MIR"

Query Match 76.8%; Score 19.2; DB 9; Length 162542;  
 Best Local Similarity 87.5%; Pred.No.1.8e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGAGACTCA 25

Db 120881 AAAAACACTTGGAGGAGACTCA 120858

## RESULT 50

AL590406

LOCUS

DEFINITION

Human DNA sequence from clone Rp11-418P12 on chromosome 6, complete

sequence.

AL590406

VERSION

AL590406.6 GI:14422236

HTG; HMGCL;

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 164831)

Smith, M.

Direct Submission

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Jun 13, 2001 this sequence version replaced gi:14270001.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Em:, ENBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the sanger Centre Chromosome 6 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>  
RP11-418P12 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOP: PRACE1.6

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

```

Location/Qualifiers
1. .164831
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="RZPD:RPCI1B753P12418"
   /db_xref="taxon:9606"
   /chromosome="6"
   /clone="RP11-418P12"
   /clone_lib="RPCI11.2"

```

misc feature

```

1 clone_1415 = RPL1-418P12"
  /notes="Clone left end: RPL1-418P12"
  join(complement(AL590290.5:2101..2.151800),complement
complement(151703)..151800),complement
complement(129720),complement
complement(129720),complement
complement(101932)..102080),complement
complement(83303)..83491),complement
complement(22254..23647)
  /locus_tag="RPL1-418P12.1-002"
  join(complement(AL590290.5:2101..2.151800),complement
complement(129613)..129720),complement
complement(101932)..102080),complement
complement(83303)..83491),complement
complement(22254..23647)

```

```

/locus_tag="RP11-418P12.1-002"
/note="match: cDNAs: AK055075.1"
join(complement(AL590230.5-2101..2247),
complement(129394..130024), complement(1
complement(104392..104487), complement(1
complement(87120..87183), complement(833
complement(27318..27443), complement(222

```

```

/locus tag="RP11-418P12.1-001"
join(complement(AL590250.5:2101..2247)
complement(139944..130024),complement(11
complement(810392..104487),complement(833
complement(87120..87183),complement(833
complement(27318..27443),complement(222
/locus tag="RP11-418P12.1-001"
/notes="match: ESTs: BG565478 BG703604 BI
match: cDNAs: AF131827.1 B1824058"

```

```
/LOCUS lag=KF11418F12.1-001
/note="match: ESTs: BG664787 BG703604 BI824058 F06678
match: cDNAs: AF131827.1 BI824058"
```

polyA site

polyA\_site  
complement(22254)  
/locus\_tag="RP11-418P12.1-001"

**polyA signal**

polyA\_signal  
complement(22276..22281)  
/locus\_tag="RP11-418P12.1-001"

CDS

	<pre>/locus_tag="RP11-418P12.1-001" /standard_name="OTTHUMP00000016648" /codon_start=1 /protein_id="CAI40662.1" /db_xref="GI:57209295" /db_xref="GOA:O5JRJ4" /db_xref="InterPro:IPR000891" /db_xref="UniProt:IREMBL:O5JRJ4" /translation="WADHTEVVKHGHPGVRYFVTPNLQGFHHVAAGATISVFG AASEFSKNIKNCSTBEESKFEVVKVSARHNNIPARGVSVSCALGPCYEGSITPQKV ETSKRLGCGCYEISLGDITGVGTGSGMRMLAESVMKEIPPGALAVHCHDTYQOALAN ILTLQMGINVDSAVSGLGGCGPYAKAGSGNVATEDLIYMLINGLGLNTGVNLYKNVMEA GPFICKAVNKTNSKVAQASRNA" join(complement(ALS590290.5:2104..&gt;2226), complement(129944..130027),complement(129613..129720), complement(99959..102080)) /locus_tag="RP11-418P12.1-003" join(complement(ALS590290.5:2104..&gt;2226), complement(129944..130027),complement(129613..129720), complement(99959..102080)) /locus_tag="RP11-418P12.1-003" /note="match: cDNAs: AL137605.1" complement(99959) /locus_tag="RP11-418P12.1-003" complement(99977..99982) /locus_tag="RP11-418P12.1-003" join(complement(ALS590290.5:2104..&gt;2226), complement(129944..130027),complement(129613..129720), complement(101922..102080)) /locus_tag="RP11-418P12.1-003" /standard_name="OTTHUMP00000016649" /note="match: proteins: Q9NT06" /codon_start=1 /protein_id="CAI40661.1" /db_xref="GI:57209294" /db_xref="GOA:Q9NT06" /db_xref="InterPro:IPR000891" /translation="AAPGAMGNVPSAVKHCLSYQQLIREHLWIGDSVAGALDPAQET SQSGHLPFKEIVGPGFQKKNKIVPTDIEKFINRLSQGLSVIEVTSVSGRW VQVAAAGATEIVFGGAASEFSKKNKINIESMGKFEVVKVSARHNNIPARGYL" 127828..127923 /note="Sequence from uni-directional primer reads and dGMP big dye terminator reads only." /notes="Clone_left_end: RP11-22806" 145034 /notes="Clone_right_end: RP11-418P12" 164831</pre>
gene	
mRNA	
polya_site	
polya_signal	
CDS	
misc_feature	
misc_feature	
misc_feature	

Query Match	76.8%	Score 19.2;	DB 8;	Length 164831;
Best Local Similarity	87.5%	Pred. No. 1.8e+02;		
Matches 21: Conservative	0: Mismatches	3: Indels	0:	

2 AAAACACTTGGAGGGGAGACTCA 25

154107 AAAAAATTGGGAAGGAAAACTCA 154130

Search completed: February 3, 2006, 21:26:27  
Job time : 876 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttgaaggagactca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	76.8	873	10	ABT42203
2	19.2	76.8	972	8	ABZ51779
3	18.6	74.4	2511	6	ABK34598
4	18.6	74.4	9210	4	AAL37257
5	18.6	74.4	9210	4	AAL37270
6	18.6	74.4	9210	8	ABX60245
7	18.6	74.4	9210	8	ABX60258
8	18.6	74.4	9210	12	ADJ30995
9	18.6	74.4	9210	12	ADJ31008
10	18.6	74.4	10818	12	ADQ97332
11	18.6	74.4	23643	12	ADQ97331_5
12	18.6	74.4	62169	12	ADQ59521
13	18.6	74.4	63529	14	ADZ13902
14	18.6	74.4	110000	14	ADZ42274_2
15	18.2	72.8	188	12	ACH99910
16	18.2	72.8	786	6	ACH76210
17	18.2	72.8	786	6	ABZ15673
18	18.2	72.8	865	11	ACN86065
19	18.2	72.8	1260	13	ADS57832

20	18.2	72.8	2092	6	ABL68555
21	18.2	72.8	2092	11	ADI31610
22	18.2	72.8	2092	13	ADS83677
23	18.2	72.8	2306	13	ADT47914
24	18.2	72.8	2307	8	ACA57281
25	18.2	72.8	6721	12	ADO31133
26	18.2	72.8	6721	13	ADZ49257
27	18.2	72.8	6787	14	ADZ49257
28	18.2	72.8	145985	12	ADQ97164
29	17.8	71.2	346	10	ADD17061
30	17.8	71.2	524	10	ADC76449
31	17.8	71.2	524	10	ADD17782
32	17.8	71.2	525	10	AKS59666
33	17.8	71.2	530	10	ADC75850
34	17.8	71.2	5347	10	ADB75375
35	17.8	71.2	30642	13	ABD33003
36	17.8	71.2	72352	12	ADQ97067
37	17.8	71.2	169998	6	ADJ36511
38	17.8	71.2	197496	6	ABN85584
39	17.8	71.2	209083	13	ABD32854
40	17.8	71.2	33462	10	ADC24763
41	17.6	70.4	120	3	AAC13345
42	17.6	70.4	299	2	AAV04596
43	17.6	70.4	299	4	AAC90858
44	17.6	70.4	300	3	AAA000406
45	17.6	70.4	389	5	AAF65920
46	17.6	70.4	398	4	AAI88689
47	17.6	70.4	436	2	AAV04595
48	17.6	70.4	436	2	AAC90876
49	17.6	70.4	436	5	AAF83431
50	17.6	70.4	480	14	ACL63156
51	17.6	70.4	555	11	ADT97028
52	17.6	70.4	555	11	ADX43510
53	17.6	70.4	563	6	ABQ58300
54	17.6	70.4	567	6	ABV88555
55	17.6	70.4	643	3	AAC06210
56	17.6	70.4	786	10	ABT41810
57	17.6	70.4	1110	6	ABL91321
58	17.6	70.4	1152	5	AAF83435
59	17.6	70.4	1152	5	AAF83436
60	17.6	70.4	1303	5	AAF83434
61	17.6	70.4	1303	5	AAF83433
62	17.6	70.4	1607	5	ABV23379
63	17.6	70.4	1966	9	ACH04153
64	17.6	70.4	2273	12	ADQ86610
65	17.6	70.4	2477	10	ADA53176
66	17.6	70.4	2516	10	ADA53332
67	17.6	70.4	2588	10	ADC30725
68	17.6	70.4	2627	11	ADI30862
69	17.6	70.4	2627	13	ADS82929
70	17.6	70.4	3760	6	ABK83731
71	17.6	70.4	3856	10	ADJ56310
72	17.6	70.4	4015	4	AAH33952
73	17.6	70.4	4029	4	AAI58323
74	17.6	70.4	4029	5	ADQ98531
75	17.6	70.4	4029	9	ADB48291
76	17.6	70.4	4331	4	AAI60109
77	17.6	70.4	5184	4	AAI03097
78	17.6	70.4	40611	12	ADQ97540
79	17.6	70.4	94752	10	ADF11646
80	17.6	70.4	110000	2	AAJ91990_10
81	17.6	70.4	110000	6	ABA90521_00
82	17.6	70.4	140036	6	AAJ98600_00
83	17.6	70.4	174742	14	ADZ13139
84	17.6	70.4	174703	11	ACN44738
85	17.6	70.4	247544	12	ADQ59419
86	17.6	70.4	247654	14	ADZ13712
87	17.6	70.4	312477	12	ADP69744
88	17.4	69.6	2682	6	ABZ14037
89	17.4	69.6	109559	13	ABD13157
90	17.2	68.8	300	2	AAZ14942
91	17.2	68.8	402	5	AAF66064
92	17.2	68.8	493	5	ABV07171

ABL68555	Kidney ca
ADI31610	Human CDN
ADS83677	Human lym
ADT47914	Bacterial
ACA57281	Human adi
ADO31133	N-end rul
ADZ49257	Pancreati
ADZ49257	Insulin s
ADQ97164	Human can
ADD17061	DNA (Segi
ADC76449	DNA homol
ADD17782	DNA (Segi
AKS59666	Plant DNA
ADC75850	DNA homol
ADB75375	Prostate
ABD33003	Mouse can
ADQ97067	Mouse can
ADJ36511	Human Her
ABN85584	Human EGF
ABD32854	Human can
ADC24763	Human wil
AAC13345	Human sec
AAV04596	Flea seri
AAC90858	Flea seri
AAA000406	Human col
AAF65920	Novel hum
AAI88689	Human pol
AAV04595	Flea seri
AAC90876	Flea seri
AAF83431	Flea seri
ACL63156	Human col
ADT97028	Colo can
ADX43510	Human CDN
ABQ58300	Human col
ABV88555	Human col
AAC06210	Human sec
ABT41810	Toxicity
ABL91321	Chlamydia
AAF83435	Flea seri
AAF83436	Flea seri
AAF83434	Flea seri
AAF83433	Flea seri
ABV23379	Human pro
ACH04153	Human CDN
ADQ86610	Human tum
ADA53176	Human cod
ADA53332	Human cod
ADC30725	Human nov
ADI30862	Human CDN
ADS82929	Human lym
ABK83731	Human CDN
ADJ56310	Human CDN
AH33952	Human col
AAI58323	Human pol
ADQ98531	DNA encod
ADB48291	Novel hum
AAI60109	Human pol
AAI03097	Human rep
ADQ97540	Mouse can
ADF11646	Human chr
AAJ91990_10	Continuation (11 o
ABA90521_00	Genomic s
AAJ98600_00	Human gen
ADZ13139	Human can
ACN44738	Human can
ADQ59419	Human can
ADZ13712	Murine ca
ADP69744	Human ROC
ABZ14037	Arabidops
ABD13157	Murine ca
AAZ14942	Human gen
AAF66064	Novel hum
ABV07171	Human pro

93	17.2	68.8	1647	12	ADN72128
94	17.2	68.8	1647	12	ADN74846
C 95	17.2	68.8	1727	8	AAS26637
C 96	17.2	68.8	1727	8	ABX73986
97	17.2	68.8	1737	10	ADB58591
98	17.2	68.8	1737	10	ADB53224
99	17.2	68.8	1737	13	ADV41392
C 100	17.2	68.8	2775	13	AQO80686
C 101	17.2	68.8	2793	6	AB212589
C 102	17.2	68.8	2793	10	AD331434
C 103	17.2	68.8	2793	12	ADI44098
104	17.2	68.8	3097	4	ABO35543
105	17.2	68.8	10368	4	ABO35200
C 106	17.2	68.8	42863	10	ADC86412
C 107	17.2	68.8	95484	12	AQO97298
C 108	17.2	68.8	110000	11	ACN44932_0
C 109	17.2	68.8	110000	14	AEA61124_1
C 110	17.2	68.0	104	9	ACH31000
C 111	17.2	68.0	112	9	ACH31003
112	17.2	68.0	201	13	ADQ41416
113	17.2	68.0	201	13	ADQ41424
114	17.2	68.0	201	13	ADQ41414
115	17.2	68.0	201	13	ADQ41426
116	17.2	68.0	387	12	ADP91769
117	17.2	68.0	396	12	ADP91770
118	17.2	68.0	441	13	ACN52012
119	17.2	68.0	467	4	AAK63939
120	17.2	68.0	492	10	ADF14124
C 121	17.2	68.0	497	9	ACH18320
C 122	17.2	68.0	524	13	ADQ54154
C 123	17.2	68.0	547	3	AC78359
C 124	17.2	68.0	553	3	ACH70372
125	17.2	68.0	586	3	AAK74925
C 126	17.2	68.0	591	10	ADD48781
127	17.2	68.0	603	13	ACN51185
C 128	17.2	68.0	604	13	ACN53437
C 129	17.2	68.0	715	10	ADC72291
C 130	17.2	68.0	982	13	ADX54334
C 131	17.2	68.0	1095	2	AAV23696
C 132	17.2	68.0	1104	5	AAS78211
C 133	17.2	68.0	1104	5	AAS75633
C 134	17.2	68.0	1104	5	AAS84590
C 135	17.2	68.0	1108	10	ADF14125
C 136	17.2	68.0	1122	12	ADM36165
C 137	17.2	68.0	1137	11	ACL28459
138	17.2	68.0	1139	4	AAK86252
139	17.2	68.0	1230	4	ABA89350
C 140	17.2	68.0	1325	13	ADR24261
C 141	17.2	68.0	1330	4	AAH34853
142	17.2	68.0	1530	14	AEA39250
C 143	17.2	68.0	1575	9	ADB09307
C 144	17.2	68.0	1575	9	ADB09301
C 145	17.2	68.0	1575	9	ADB09309
C 146	17.2	68.0	1575	9	ADB09303
C 147	17.2	68.0	1575	9	ADB09305
C 148	17.2	68.0	1636	6	ABK34741
C 149	17.2	68.0	1656	2	AAK60810
C 150	17.2	68.0	1656	4	AA59216

## ALIGNMENTS

RESULT 1

ABT42203

ID ABT42203 standard; DNA; 873 BP.

XX

AC ABT42203;

XX

DT 26-JUN-2003 (first entry)

XX

DE Toxicity modelling related rat gene SEQ ID No 1905.

XX

Adn72128 Thale cre  
Adn74846 Thale cre  
Aas26637 Human gen  
Abx73986 Human nov  
Adb58591 Toxicity-  
Adb53224 Primary r  
Adv41392 Rat cardi  
Adq80686 Arabidops  
Ab212589 Arabidops  
Ade31434 Plant yie  
Adi44098 Plant tra  
Abo35543 Novel mou  
Ab105200 Drosophil  
Adc86412 Human GPC  
Adq97298 Mouse can  
Acn44932 Mouse gen  
Continuation (2 of  
Ach31000 Human bon  
Ach31003 Human bon  
Adq41416 Myocardia  
Adq41424 Myocardia  
Adq41414 Myocardia  
Adq41426 Myocardia  
Adp91769 Cotton ex  
Adp91770 Cotton ex  
Acs2012 Cotton an  
Aak63939 Human inm  
Adf14124 Human end  
Ach18320 Human adu  
Adg54154 Novel can  
Aac78359 Human can  
Ach70372 Human gen  
Aac74925 Human ORF  
Add48781 Rat gene  
Acs51185 Cotton an  
Acs53437 Cotton an  
Adc72291 Poppy DNA  
Adx54334 Plant ful  
Aav23696 Human NSP  
Aas78211 DNA encod  
Aas75633 DNA encod  
Aas84590 DNA encod  
Adf14125 Human end  
Adm36165 Human RTN  
Acl28459 Rice abio  
Aak86252 Human inm  
Aba89350 Escherich  
Adr24261 Breast ca  
Aah34853 Human col  
Aea39250 Zizania (  
Adb09307 Alloioococ  
Adb09301 Alloioococ  
Adb09309 Alloioococ  
Adb09303 Alloioococ  
Adb09305 Alloioococ  
Adbk34741 Human CDN  
Aax60810 Human sec  
Aas59216 Human CDN

KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;  
database; drug screening; toxicity assay; rat; ds.

XX Rattus norvegicus.

PN WO200295000-A2.

XX 28-NOV-2002.

PD 22-MAY-2002; 2002WO-US016173.

XX 22-MAY-2001; 2001US-0292335P.

PR 13-JUN-2001; 2001US-0297523P.

PR 19-JUN-2001; 2001US-0298925P.

PR 10-JUL-2001; 2001US-0303807P.

PR 10-JUL-2001; 2001US-0303808P.

PR 10-JUL-2001; 2001US-0303810P.

PR 28-AUG-2001; 2001US-0315047P.

PR 27-SEP-2001; 2001US-0324928P.

PR 22-OCT-2001; 2001US-0330462P.

PR 01-NOV-2001; 2001US-0330867P.

PR 21-NOV-2001; 2001US-0331805P.

PR 06-DEC-2001; 2001US-0336144P.

PR 19-DEC-2001; 2001US-0340873P.

PR 21-FEB-2002; 2002US-0357842P.

PR 21-FEB-2002; 2002US-0357843P.

PR 21-FEB-2002; 2002US-0357844P.

PR 15-MAR-2002; 2002US-0364134P.

PR 08-APR-2002; 2002US-0370144P.

PR 08-APR-2002; 2002US-0370206P.

PR 08-APR-2002; 2002US-0370247P.

PR 17-APR-2002; 2002US-0372794P.

PR 21-APR-2002; 2002US-0371679P.

(GENE-) GENE LOGIC INC.

Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

WPI; 2003-148464/14.

Predicting at least one toxic effect of a compound, useful for toxicity

modeling, comprises preparing a gene expression profile of a tissue or

cell sample exposed to the compound, and comparing the gene expression

profile to a database.

Example 4; Page: 446pp; English.

The invention relates to a novel method of predicting at least one toxic

effect of a compound. The method comprises a gene expression profile of a

tissue or cell sample exposed to the compound, and comparing the gene

expression profile to a database comprising at least part of the data or

information given in the specification. The methods are useful for

predicting at least one toxic effect of a compound, predicting the renal

progression of a toxic effect of a compound, predicting the renal

toxicity of a compound, or identifying toxicity markers in tissues or

cells exposed to known renal toxin. The genes are useful as toxicity

markers in drug screening and toxicity assays, in monitoring disease or

physiological states, or disease progression. This polynucleotide

represents a rat DNA sequence relating to the toxic effect database

described in the specification. NOTE: The sequence data for this patent

did not form part of the printed specification, but was obtained in

electronic format directly from the World Intellectual Property

Organization

Sequence 873 BP; 213 A; 245 C; 225 G; 190 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 873;

Best Local Similarity 87.5%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25

|||||

658 AAGACACTTGCAGGGTGACTCA 681

Db





XX Human musculoskeletal system related polynucleotide SEQ ID NO 3622.  
DE  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155367-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001338.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-02556719P.

```
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal
PT cancers and also for testing and detection e.g. diagnosis.
XX
XX Example 2; SEQ ID NO 3622; 781pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 9210 BP; 2878 A; 1772 C; 1716 G; 2844 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 4; Length 9210;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACACTTGGGAAGGAGACTCA 25
Db 7719 AAACACATTGGGAAGGAGATTCA 7695
||||| ||||| ||||| ||||| |||||
RESULT 5
AAL37270/c
ID AAL37270 Standard; DNA; 9210 BP.
XX
AC AAL37270;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3635.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
XX
XX Homo sapiens.
XX
XX WO200155367-A1.
XX
XX 02-AUG-2001.
XX
XX
```

```
XX 17-JAN-2001; 2001WO-US001338.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 21-SEP-2000; 2000US-0233365P.
XX 21-SEP-2000; 2000US-0234223P.
XX 25-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234997P.
XX 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
XX 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX
```

29-SEP-2000; 2000US-0236369P.  
 29-SEP-2000; 2000US-0236370P.  
 02-OCT-2000; 2000US-0236802P.  
 02-OCT-2000; 2000US-0237037P.  
 02-OCT-2000; 2000US-0237038P.  
 02-OCT-2000; 2000US-0237039P.  
 02-OCT-2000; 2000US-0237040P.  
 02-OCT-2000; 2000US-0239935P.  
 13-OCT-2000; 2000US-0239937P.  
 20-OCT-2000; 2000US-0240960P.  
 20-OCT-2000; 2000US-0241221P.  
 20-OCT-2000; 2000US-0241785P.  
 20-OCT-2000; 2000US-0241786P.  
 20-OCT-2000; 2000US-0241787P.  
 20-OCT-2000; 2000US-0241808P.  
 20-OCT-2000; 2000US-0241826P.  
 01-NOV-2000; 2000US-0244617P.  
 08-NOV-2000; 2000US-0246474P.  
 08-NOV-2000; 2000US-0246475P.  
 08-NOV-2000; 2000US-0246476P.  
 08-NOV-2000; 2000US-0246478P.  
 08-NOV-2000; 2000US-0246523P.  
 08-NOV-2000; 2000US-0246524P.  
 08-NOV-2000; 2000US-0246525P.  
 08-NOV-2000; 2000US-0246526P.  
 08-NOV-2000; 2000US-0246527P.  
 08-NOV-2000; 2000US-0246528P.  
 08-NOV-2000; 2000US-0246532P.  
 08-NOV-2000; 2000US-0246609P.  
 08-NOV-2000; 2000US-0246610P.  
 08-NOV-2000; 2000US-0246611P.  
 08-NOV-2000; 2000US-0246613P.  
 17-NOV-2000; 2000US-0249207P.  
 17-NOV-2000; 2000US-0249208P.  
 17-NOV-2000; 2000US-0249209P.  
 17-NOV-2000; 2000US-0249210P.  
 17-NOV-2000; 2000US-0249211P.  
 17-NOV-2000; 2000US-0249212P.  
 17-NOV-2000; 2000US-0249213P.  
 17-NOV-2000; 2000US-0249214P.  
 17-NOV-2000; 2000US-0249215P.  
 17-NOV-2000; 2000US-0249216P.  
 17-NOV-2000; 2000US-0249217P.  
 17-NOV-2000; 2000US-0249218P.  
 17-NOV-2000; 2000US-0249244P.  
 17-NOV-2000; 2000US-0249245P.  
 17-NOV-2000; 2000US-0249264P.  
 17-NOV-2000; 2000US-0249265P.  
 17-NOV-2000; 2000US-0249297P.  
 17-NOV-2000; 2000US-0249299P.  
 17-NOV-2000; 2000US-0249300P.  
 01-DEC-2000; 2000US-0250160P.  
 01-DEC-2000; 2000US-0250391P.  
 05-DEC-2000; 2000US-0251030P.  
 05-DEC-2000; 2000US-0251988P.  
 05-DEC-2000; 2000US-0256719P.  
 06-DEC-2000; 2000US-0251479P.  
 08-DEC-2000; 2000US-0251856P.  
 08-DEC-2000; 2000US-0251868P.  
 08-DEC-2000; 2000US-0251869P.  
 08-DEC-2000; 2000US-0251989P.  
 08-DEC-2000; 2000US-0251990P.  
 11-DEC-2000; 2000US-0254097P.  
 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-451937/48.

Isolated polypeptide for treating, preventing and/ or prognosing disorders related to the musculoskeletal system including musculoskeletal cancers and also for testing and detection e.g. diagnosis.  
 Example 2; SEQ ID NO 3635; 781pp + Sequence Listing; English.  
 The invention relates to novel genes (AAL34669-AAL37666) and proteins (ABR03087-ABR04109) associated with the musculoskeletal system useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIFO at ftp.wipo.int/pub/published\_pct\_sequences  
 Query Match 74.4%; Score 18.6; DB 4; Length 9210;  
 Best Local Similarity 84.0%; Pred. No. 2.7e-02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAACAACCTTGGAGGAGGACTCA 25  
 Db 7719 AAAACACATTGGAGGAGGATTCA 7695  
 RESULT 6  
 ABX60245/c  
 ID ABX60245 standard; cDNA; 9210 BP.  
 AC ABX60245;  
 XX  
 XX 26-FEB-2003 (first entry)  
 XX  
 DE cDNA encoding novel human musculoskeletal system antigen #2589.  
 XX  
 KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
 KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
 KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
 KW post-operative tissue repair; limb regeneration; neuronal growth;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW AIDS-related complex; chondrocyte growth; bone regeneration;  
 KW periodontal regeneration; tissue transport; bone graft; skin aging;  
 KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
 KW cell growth; organ transplant; cell differentiation; body height; weight;  
 KW hair colour; eye colour; skin; percentage of adipose tissue;  
 KW pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;  
 KW depression; tendency for violence; pain; reproductive capability;  
 KW hormone level; endocrine level; appetite; libido; memory; stress;  
 KW storage capability; fat content; lipid content; protein content;  
 KW carbohydrate content; vitamin content; cofactor content;  
 KW nutritional component.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US2002147140-A1.  
 XX  
 XX 10-OCT-2002.  
 PD  
 XX  
 XX 17-JAN-2001; 2001US-00764877.  
 PF  
 XX 31-JAN-2000; 2000US-0179065P.  
 XX  
 XX 04-FEB-2000; 2000US-0180628P.  
 PR

PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229133P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 23-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA.; Ruben SM, Barash SC;  
 DR WPT; 2003-128199/12.  
 XX  
 PT Isolated nucleic acid molecules encoding musculoskeletal system  
 PT associated polypeptides, useful for detecting disorders, e.g. cancer.  
 XX  
 PS Disclosure; SEQ ID NO 3622; 321pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule comprising a  
 CC sequence encoding musculoskeletal system associated polypeptides useful  
 CC for detecting disorders, e.g., cancer or cancer metastases, in animals or  
 CC humans. The nucleic acid; stimulates re-vascularisation of ischaemic  
 CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
 CC and other cardiovascular conditions; treats wounds due to injuries,  
 CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
 CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
 CC neuronal damage occurring in certain disorders or neurodegenerative  
 CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
 CC related complex; stimulates chondrocyte growth, thus they can be used to

CC enhance bone and periodontal regeneration and aid in tissue transports or  
 CC bone grafts; prevents skin aging due to sunburn by stimulating  
 CC keratinocyte growth; prevents hair loss, since FGF family members  
 CC activate hair-forming cells and promotes melanocyte growth; stimulates  
 CC growth and differentiation of hematopoietic cells and bone marrow cells  
 CC when used in combination with other cytokines; maintains organs before  
 CC transplantation or for supporting cell culture of primary tissues;  
 CC induces tissue of mesodermal origin to differentiate in early embryos;  
 CC increases or decreases the differentiation or proliferation of embryonic  
 CC stem cells, besides, haematopoietic lineage; modulates mammalian  
 CC characteristics, such as, body height, weight, hair colour, eye colour,  
 CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
 CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
 CC state or physical state by influencing biorhythms, cardiac rhythms,  
 CC depression, tendency for violence, tolerance for pain, reproductive  
 CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
 CC stress; increases or decreases storage capabilities, fat content, lipid,  
 CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
 CC components. This sequence encodes a novel human musculoskeletal system  
 CC antigen. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from the US patent office at  
 CC ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140  
 CC  
 XX Sequence 9210 BP; 2878 A; 1772 C; 1716 G; 2844 T; 0 U; 0 Other;  
 SQ  
 Query Match 74.4%; Score 18.6; DB 8; Length 9210;  
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACACTTGGAGGGAGACTCA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 7719 AAAACACATTGGAGGAGATTCA 7695  
 RESULT 7  
 ABX60258/C  
 ID ABX60258 standard; cDNA; 9210 BP.  
 XX  
 AC ABX60258;  
 XX  
 XX 26-FEB-2003 (first entry)  
 DT  
 XX  
 DE cDNA encoding novel human musculoskeletal system antigen #2602.  
 XX  
 KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
 KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
 KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
 KW post-operative tissue repair; limb regeneration; neuronal growth;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW AIDS-related complex; chondrocyte growth; bone regeneration;  
 KW periodontal regeneration; tissue transport; bone graft; skin aging;  
 KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
 KW cell growth; organ transplant; cell differentiation; body height; weight;  
 KW hair colour; eye colour; skin; percentage of adipose tissue;  
 KW pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;  
 KW depression; tendency for violence; pain; reproductive capability;  
 KW hormone level; endocrine level; appetite; libido; memory; stress;  
 KW storage capability; fat content; lipid content; protein content;  
 KW carbohydrate content; vitamin content; cofactor content;  
 KW nutritional component.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2002147140-A1.  
 PN  
 XX 10-OCT-2002.  
 PD  
 XX 17-JAN-2001; 2001US-00764877.  
 PF  
 XX 31-JAN-2000; 2000US-0179065P.  
 PR  
 XX 04-FEB-2000; 2000US-0180628P.  
 PR  
 XX 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-128199/12.  
XX  
PT Isolated nucleic acid molecules encoding musculoskeletal system  
PT associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
PS Disclosure; SEQ ID NO 3635; 321pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule comprising a  
CC sequence encoding musculoskeletal system associated polypeptides useful  
CC for detecting disorders, e.g., cancer or cancer metastases, in animals or  
CC humans. The nucleic acid; stimulates re-vascularisation of ischaemic  
CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
CC and other cardiovascular conditions; treats wounds due to injuries,  
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or

CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, such as FGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, hematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, eye colour,  
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, circadian rhythms,  
CC depression, tendency for violence, tolerance for pain, reproductive  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140  
XX  
SQ Sequence 9210 BP; 2878 A; 1772 C; 1716 G; 2844 T; 0 U; 0 Other;  
Query Match 74.4%; Score 18.6; DB 8; Length 9210;  
Best Local Similarity 84.0%; Pred. No. 2.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGGAGACTCA 25  
Db 7719 AAACACATTGGAGGAAGATTCA 7695  
||||| ||||| ||||| ||||| |||||  
RESULT 8  
ADJ30995/c  
ID ADJ30995 standard; DNA; 9210 BP.  
XX  
AC ADJ30995;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE Human musculoskeletal system-associated genomic DNA - SEQ ID 3622.  
XX  
KW musculoskeletal system; cytostatic; osteopathic; cancer; osteoporosis;  
KW gene therapy; vaccine; human; ds.  
XX  
OS Homo sapiens.  
XX  
PN US2004009488-A1.  
XX  
PD 15-JAN-2004.  
XX  
PF 13-SEP-2002; 2002US-00242515.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.





Best Local Similarity 84.0%; Pred. No. 2.7e+02;			
Matches	21; Conservative	0; Mismatches	4; Indels
		0; Gaps	0;
QY	1 AAAAAACACTTGGAAAGGAGACTCA 25		
Db	7719 AAAACACATTGGAAGGAGATTCA 7695		
RESULT 9			
ADJ31008/c			
ID	ADJ31008 standard; DNA; 9210 BP.		
XX			
AC	ADJ31008;		
XX			
DT	20-MAY-2004 (first entry)		
XX			
DE	Human musculoskeletal system-associated genomic DNA - SEQ ID 3635.		
XX			
KW	musculoskeletal system; cytostatic; osteopathic; cancer; osteoporosis;		
KW	gene therapy; vaccine; human; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	US2004009488-A1.		
XX			
PD	15-JAN-2004.		
XX			
PF	13-SEP-2002; 2002US-00242515.		
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-0205515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	07-JUL-2000; 2000US-0216880P.		
PR	11-JUL-2000; 2000US-0217487P.		
PR	11-JUL-2000; 2000US-0217496P.		
PR	14-JUL-2000; 2000US-0218290P.		
PR	26-JUL-2000; 2000US-0220963P.		
PR	26-JUL-2000; 2000US-0220964P.		
PR	14-AUG-2000; 2000US-0224518P.		
PR	14-AUG-2000; 2000US-0224519P.		
PR	14-AUG-2000; 2000US-0225213P.		
PR	14-AUG-2000; 2000US-0225214P.		
PR	14-AUG-2000; 2000US-0225266P.		
PR	14-AUG-2000; 2000US-0225267P.		
PR	14-AUG-2000; 2000US-0225268P.		
PR	14-AUG-2000; 2000US-0225270P.		
PR	14-AUG-2000; 2000US-0225447P.		
PR	14-AUG-2000; 2000US-0225757P.		
PR	14-AUG-2000; 2000US-0225758P.		
PR	14-AUG-2000; 2000US-0225759P.		
PR	18-AUG-2000; 2000US-0226279P.		
PR	22-AUG-2000; 2000US-0226681P.		
PR	22-AUG-2000; 2000US-0226868P.		
PR	23-AUG-2000; 2000US-0227182P.		
PR	23-AUG-2000; 2000US-0227009P.		
PR	30-AUG-2000; 2000US-0228924P.		
PR	01-SEP-2000; 2000US-0229287P.		
PR	01-SEP-2000; 2000US-0229343P.		
PR	01-SEP-2000; 2000US-0229344P.		
PR	01-SEP-2000; 2000US-0229345P.		
PR	05-SEP-2000; 2000US-0229509P.		
PR	05-SEP-2000; 2000US-0229513P.		
PR	06-SEP-2000; 2000US-0230437P.		
PR	06-SEP-2000; 2000US-0230438P.		

PR	08-SEP-2000; 2000US-0231242P.
PR	08-SEP-2000; 2000US-0231243P.
PR	08-SEP-2000; 2000US-0231244P.
PR	08-SEP-2000; 2000US-0231413P.
PR	08-SEP-2000; 2000US-0231414P.
PR	08-SEP-2000; 2000US-0232080P.
PR	08-SEP-2000; 2000US-0232081P.
PR	12-SEP-2000; 2000US-0231968P.
PR	14-SEP-2000; 2000US-0232397P.
PR	14-SEP-2000; 2000US-0232398P.
PR	14-SEP-2000; 2000US-0232399P.
PR	14-SEP-2000; 2000US-0232400P.
PR	14-SEP-2000; 2000US-0232401P.
PR	14-SEP-2000; 2000US-0233063P.
PR	14-SEP-2000; 2000US-0233064P.
PR	14-SEP-2000; 2000US-0233065P.
PR	21-SEP-2000; 2000US-0234223P.
PR	21-SEP-2000; 2000US-0234274P.
PR	25-SEP-2000; 2000US-0234997P.
PR	25-SEP-2000; 2000US-0234998P.
PR	26-SEP-2000; 2000US-0235484P.
PR	27-SEP-2000; 2000US-0235834P.
PR	29-SEP-2000; 2000US-0236327P.
PR	29-SEP-2000; 2000US-0236367P.
PR	29-SEP-2000; 2000US-0236368P.
PR	29-SEP-2000; 2000US-0236369P.
PR	29-SEP-2000; 2000US-0236370P.
PR	02-OCT-2000; 2000US-0237032P.
PR	02-OCT-2000; 2000US-0237037P.
PR	02-OCT-2000; 2000US-0237038P.
PR	02-OCT-2000; 2000US-0237039P.
PR	02-OCT-2000; 2000US-0237040P.
PR	13-OCT-2000; 2000US-0239933P.
PR	13-OCT-2000; 2000US-0239937P.
PR	20-OCT-2000; 2000US-0240960P.
PR	20-OCT-2000; 2000US-0241221P.
PR	20-OCT-2000; 2000US-0241785P.
PR	20-OCT-2000; 2000US-0241786P.
PR	20-OCT-2000; 2000US-0241787P.
PR	20-OCT-2000; 2000US-0241808P.
PR	20-OCT-2000; 2000US-0241809P.
PR	20-OCT-2000; 2000US-0241826P.
PR	01-NOV-2000; 2000US-0244617P.
PR	08-NOV-2000; 2000US-0244647P.
PR	08-NOV-2000; 2000US-0246475P.
PR	08-NOV-2000; 2000US-0246476P.
PR	08-NOV-2000; 2000US-0246477P.
PR	08-NOV-2000; 2000US-0246478P.
PR	08-NOV-2000; 2000US-0246523P.
PR	08-NOV-2000; 2000US-0246524P.
PR	08-NOV-2000; 2000US-0246525P.
PR	08-NOV-2000; 2000US-0246526P.
PR	08-NOV-2000; 2000US-0246527P.
PR	08-NOV-2000; 2000US-0246528P.
PR	08-NOV-2000; 2000US-0246532P.
PR	08-NOV-2000; 2000US-0246609P.
PR	08-NOV-2000; 2000US-0246610P.
PR	08-NOV-2000; 2000US-0246611P.
PR	08-NOV-2000; 2000US-0246613P.
PR	17-NOV-2000; 2000US-0249207P.
PR	17-NOV-2000; 2000US-0249208P.
PR	17-NOV-2000; 2000US-0249210P.
PR	17-NOV-2000; 2000US-0249211P.
PR	17-NOV-2000; 2000US-0249212P.
PR	17-NOV-2000; 2000US-0249213P.
PR	17-NOV-2000; 2000US-0249214P.
PR	17-NOV-2000; 2000US-0249215P.
PR	17-NOV-2000; 2000US-0249216P.
PR	17-NOV-2000; 2000US-0249217P.
PR	17-NOV-2000; 2000US-0249218P.
PR	17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251030P.  
PR 03-DEC-2000; 2000US-0251988P.  
PR 03-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764877.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI; 2004-090458/09.  
DR  
XX  
XX New nucleic acid molecule, useful for preparing a medicament for  
PT preventing, treating or ameliorating a medical condition e.g., cancer of  
PT musculoskeletal tissues or osteoporosis.  
XX  
XX  
XX Disclosure; SEQ ID NO 3635; 289pp; English.  
PS  
XX  
XX The invention relates to a novel isolated musculoskeletal system-  
CC associated nucleic acid molecule. The nucleic acid of the invention  
CC demonstrates cytostatic and osteopathic activities and may be useful for  
CC preparing a medicament for preventing, treating or ameliorating a medical  
CC condition such as cancer of the musculoskeletal tissues or osteoporosis,  
CC possibly via gene therapy or vaccine production. The current sequence is  
CC that of the human musculoskeletal system-associated genomic DNA of the  
CC invention. The current sequence is not shown within the specification per  
CC se but is available on the USPTO web-site  
CC http:seqdata.uspto.gov/sequence.html?DocID=20040009488.  
XX  
XX Sequence 9210 BP; 2878 A; 1772 C; 1716 G; 2844 T; 0 U; 0 Other;  
SQ  
Query Match 74.4%; Score 18.6; DB 12; Length 9210;  
Best Local Similarity 84.0%; Pred. No. 2.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGAGACTCA 25  
Db 7719 AAAAAACACTTGGAGGAGACTCA 7695  
RESULT 10  
ADQ97332/c  
ID ADQ97332 standard; DNA; 10818 BP.  
XX  
XX AC  
XX ADQ97332;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
XX Human cancer associated sequence HR08-032, SEQ ID 309.  
XX  
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO2004060304-A2.  
PN  
XX  
XX 22-JUL-2004.  
PD  
XX

PF 22-DEC-2003; 2003WO-US041389.  
XX  
XX 27-DEC-2002; 2002US-00330773.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
PA  
XX  
XX Morris DW, Malandro MS;  
PI  
XX WPI; 2004-543781/52.  
DR  
XX  
XX New isolated cancer associated nucleic acids comprising at least 10  
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
PT cancers such as leukemia and lymphoma.  
XX  
XX  
XX Claim 1; SEQ ID NO 309; 199pp; English.  
PS  
XX  
XX The present invention relates to cancer associated sequences (ADQ97025-  
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
XX Sequence 10818 BP; 3402 A; 2198 C; 2281 G; 2937 T; 0 U; 0 Other;  
SQ  
Query Match 74.4%; Score 18.6; DB 12; Length 10818;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGAGACTCA 25  
Db 7262 AAAAAACACTTGGAGGAGACTCA 7238  
RESULT 11  
ADQ97331\_5/c  
Continuation (6 of 6) of ADQ97331 from base 500001 (Human cancer associated sequence HD  
WP Sequence split into 6 fragments LOCUS ADQ97331 Accession Adq97331  
WP Fragment Name Begin End  
WP ADQ97331\_0 1 110000  
WP ADQ97331\_1 100001 210000  
WP ADQ97331\_2 200001 310000  
WP ADQ97331\_3 300001 410000  
WP ADQ97331\_4 400001 510000  
WP ADQ97331\_5 500001 523643  
Query Match 74.4%; Score 18.6; DB 12; Length 23643;  
Best Local Similarity 84.0%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGAGACTCA 25  
Db 10087 AAAAAACACTTGGAGGAGACTCA 10063  
RESULT 12  
ADQ59521/c  
ID ADQ59521 standard; DNA; 62169 BP.  
XX  
XX AC  
XX ADQ59521;  
XX  
XX 07-OCT-2004 (first entry)  
DT  
XX  
XX Human cancer-associated (CA) gene sequence SEQ ID NO:157.  
DE  
XX  
XX human; cancer-associated gene; cancer-associated protein; cytostatic;  
KW gene therapy; vaccine; tyrosine kinase antagonist;  
KW G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO2004058288-A1.  
PN  
XX

PD 15-JUL-2004.  
XX  
XX  
PF 15-DEC-2003; 2003WO-US040082.  
XX  
XX 17-DEC-2002; 2002US-00322696.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX Morris DW, Malandro MS;  
XX  
XX WPI; 2004-543349/52.  
DR P-PSDB; ADQ59523.  
XX  
XX New cancer-associated nucleic acid for diagnosing, preventing or treating  
PT cancer (e.g. lymphoma) or for screening agents that may be used for  
PT treating or preventing cancer.  
XX  
XX Disclosure; SEQ ID NO 157; 143pp; English.  
XX  
XX The present invention describes human cancer-associated (CA) nucleotide  
CC sequences (I). Also described: (1) an expression vector comprising (I);  
CC (2) a host cell comprising (I) or the expression vector; (3) a microarray  
CC for detecting a CA nucleic acid; (4) an isolated polypeptide encoded  
CC within an open reading frame of a CA sequence; (5) an isolated antibody,  
CC or its antigen binding fragment, that binds to the above polypeptide; (6)  
CC a hybridoma that produces the monoclonal antibody described above; (7) a  
CC pharmaceutical composition comprising the antibody and a pharmaceutical  
CC excipient; (8) a kit for detecting or diagnosing cancer cells, comprising  
CC the above (monoclonal) antibody or polynucleotide that selectively  
CC hybridizes to any of the polynucleotide sequences mentioned above; (9)  
CC methods for diagnosing cancer or for detecting the presence or absence of  
CC cancer cells in an individual; (10) a method for inhibiting growth of  
CC cancer cells in an individual; (11) a method for delivering a therapeutic  
CC agent to cancer cells in an individual; (12) an electronic library  
CC comprising the polynucleotide or polypeptide, or their fragments,  
CC mentioned above; (13) a method of screening for anticancer activity; (14)  
CC methods for detecting cancer associated with expression of a polypeptide  
CC or the presence of the antibody in a test cell or serum sample; (15) a  
CC method for screening for a bioactive agent capable of modulating the  
CC activity of a CA protein encoded by the above nucleic acid molecule; and  
CC (16) a method for treating cancers. (I) has cytostatic activity, and can  
CC be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,  
CC and as a G-protein coupled receptor antagonist. The compositions and  
CC methods of the present invention can be used for diagnosing, preventing  
CC and treating cancer, especially lymphomas. They may also be used in  
CC screening for agents that may be used for treating or preventing cancer.  
CC The present sequence represents a human CA gene sequence, which is given  
CC in the exemplification of the present invention. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 62169 BP; 17028 A; 13404 C; 13215 G; 18292 T; 0 U; 230 Other;  
SQ  
Query Match 74.4%; Score 18.6; DB 12; Length 62169;  
Best Local Similarity 84.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGAAGGGAGACTCA 25  
DB 7740 AAAGGAACCTTGAAGGGTGACTCA 7716  
RESULT 13  
ADZ13902/C  
ID ADZ13902 standard; DNA; 63529 BP.  
XX  
XX ADZ13902;  
AC  
XX  
XX 16-JUN-2005 (first entry)  
DT  
XX  
XX Murine cancer-associated genomic DNA #124.  
DE  
XX

KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;  
KW cytostatic; gene; ds.  
XX  
XX Mus sp.  
XX WO2005031001-A2.  
XX  
XX 07-APR-2005.  
XX  
XX 23-SEP-2004; 2004WO-US031617.  
PF  
XX 23-SEP-2003; 2003US-00669920.  
XX  
XX (CHIR ) CHIRON CORP.  
PA  
XX Morris DW, Malandro MS;  
PI  
XX WPI; 2005-273395/28.  
DR  
XX Nucleic acid array useful for detecting cancer associated nucleic acid,  
PT comprises two or more nucleic acid probes.  
PT  
XX Disclosure; SEQ ID NO 1422; 199pp; English.  
XX  
XX The invention relates to a nucleic acid array for detecting a cancer  
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
CC The invention also relates to a peptide array comprising two or more  
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
CC that binds to a polypeptide, an isolated antibody or its fragment which  
CC binds to a polypeptide, which is prepared by immunizing a host animal  
CC with a composition comprising the polypeptide or its antigen binding  
CC fragment and collecting cells from the host expressing antibodies against  
CC the antigen or its antigen binding fragment, a composition comprising the  
CC antibody and a carrier, a method of screening for anticancer activity, a  
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
CC method of treating cancer and a method of inhibiting expression of a CA  
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
CC nucleic acids. The antibody is useful for detecting the presence or  
CC absence of cancer cells in an individual which involves contacting cells  
CC from the individual with the antibody and detecting a complex of a CA  
CC protein from the cancer cells and the antibody, where the detection of  
CC the complex correlates with the presence of cancer cells in the  
CC individual. The composition is useful for inhibiting growth of cancer  
CC cells in an individual or for delivering a therapeutic agent to cancer  
CC cells in an individual. The invention is also useful for diagnosing  
CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
CC a cell. This sequence represents murine cancer-associated genomic DNA of  
CC the invention.  
XX  
XX Sequence 63529 BP; 17363 A; 13779 C; 13492 G; 18665 T; 0 U; 230 Other;  
SQ  
Query Match 74.4%; Score 18.6; DB 14; Length 63529;  
Best Local Similarity 84.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGAAGGGAGACTCA 25  
DB 7763 AAAGGAACCTTGAAGGGTGACTCA 7739  
RESULT 14  
ADZ42274\_2  
Continuation (3 of 4) of ADZ42274 from base 200001 (Human mineralocorticoid receptor ge  
WP Sequence split into 4 fragments LOCUS ADZ42274 Accession Adz42274  
Fragment Name Begin End  
WP ADZ42274\_0 1 110000  
WP ADZ42274\_1 100001 210000  
WP ADZ42274\_2 200001 310000  
WP ADZ42274\_3 300001 369000  
Query Match 74.4%; Score 18.6; DB 14; Length 110000;  
Best Local Similarity 84.0%; Pred. No. 3.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  
1 AAAAAACACTTGAAGGGGAGACTCA 25

Dp  
45145 AGAAGACACTTCGAAGGGGAGAATCA 45169

RESULT 15	
ACH89910/c	
ID	ACH89910 standard; DNA; 188 BP.
AC	
XX	ACH89910;
XX	
XX	29-JUL-2004 (first entry)
DT	
DT	
XX	
XX	Human genome derived single exon probe #23105.
DE	
XX	
XX	Human; probe; ss; gene expression; single exon probe; microarray;
KW	alternative splicing event; genomic alteration.
KW	
XX	
XX	Homo sapiens.
OS	
XX	
XX	US2003194704-A1.
PN	
XX	
XX	16-OCT-2003.
PD	
XX	
XX	03-APR-2002; 2002US-00029386.
PF	
XX	
XX	03-APR-2002; 2002US-00029386.
PR	
XX	
XX	(PENN/) PENN S G.
PA	
PA	(RANK/) RANK D R.
PA	(HANZ/) HANZEL D K.
XX	
XX	
PI	Penn SG, Rank DR, Hanzel DK;
XX	
XX	WPI; 2004-119264/12.
DR	
XX	
XX	New human genome-derived single exon nucleic acid probes useful for
PT	gene expression analysis, for identifying or characterizing altera
PT	splicing events, for assessing genomic alterations or as tools fo
PT	surveying tissues.
PT	

CC	smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030194704
XX	Sequence 188 BP; 54 A; 34 C; 23 G; 77 T; 0 U; 0 Other;
QQ	Query Match 72.8%; Score 18.2; DB 12; Length 188; Best Local Similarity 87.0%; Pred. No. 2.4e+02; Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1 AAAAAACACTTGGGAAGGAGACT 23 
DB	86 AAACACACTTGGCGAAGGAGACT 64
RESULT 16	
ACH76210/c	
ID	ACH76210 standard; DNA; 590 BP.
XX	
AC	ACH76210;
XX	
DT	29-JUL-2004 (first entry)
XX	
DE	Human genome derived single exon probe #9405.
XX	
XX	Human; probe; es; gene expression; single exon probe; microarray; alternative splicing event; genomic alteration.
KW	
KW	
OS	Homo sapiens.
XX	
PN	US2003194704-A1.
XX	
FD	16-OCT-2003.
XX	
PF	03-APR-2002; 2002US-00029386.
XX	
PR	03-APR-2002; 2002US-00029386.
XX	
PA	(PENN/) PENN S G.
PA	(RANK/) RANK D R.
PA	(HANZ/) HANZEL D K.
XX	
PI	Penn SG, Rank DR, Hanzel DK;
XX	
DR	WPI, 2004-119264/12.
XX	
PT	New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for surveying tissues.
PT	
XX	Claim 15; SEQ ID NO 9405; 80pp; English.
FS	
XX	
CC	The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above,

CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subcription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC -expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterising  
CC alternative splicing events, in detecting and characterising gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
SQ Sequence 590 BP; 193 A; 112 C; 81 G; 204 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 12; Length 590;  
Best Local Similarity 87.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACAACCTTGGAGGGAGACT 23  
||| ||||| ||||| ||||| |||||  
DB 275 AAACAACAACCTTGGCAAGGAGACT 253

RESULT 17  
ABZ15673  
ID ABZ15673 standard; DNA; 786 BP.  
XX AC ABZ15673;  
XX 21-JAN-2003 (first entry)  
DT  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 3478.  
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
KW Arabidopsis thaliana.  
XX Arabidopsis thaliana.  
OS  
XX WO200216655-A2.  
PN  
XX 28-FEB-2002.  
PD  
XX 24-AUG-2001; 2001WO-US026685.  
PF  
XX 24-AUG-2000; 2000US-0227866P.  
PR 26-JAN-2001; 2001US-0264647P.  
PR 22-JUN-2001; 2001US-0300111P.  
XX  
XX (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
DR  
XX

XX Identifying a stress condition to which a plant cell has been exposed and  
PT producing plants with increased tolerance to these abiotic stresses.  
XX  
XX Claim 144; SEQ ID NO 3478; 577pp + Sequence Listing; English.  
PS  
XX The invention relates to identifying a stress condition to which a plant  
XX cell has been exposed, comprising: (a) contacting nucleic acid  
CC representative of expressed polynucleotides in the plant cell with an  
CC array or probes representative of the plant cell genome; and (b)  
CC detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants  
XX

CC with increased tolerance to abiotic stress. The present sequence is that  
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
CC in methods of the invention. Note: The sequence data for this patent is  
CC not represented in the printed specification but is based on sequence  
CC information supplied to Derwent by the European Patent Office  
XX  
SQ Sequence 786 BP; 248 A; 147 C; 118 G; 273 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 6; Length 786;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACAACCTTGGAGGGAGACTC 24  
||| ||||| ||||| ||||| |||||  
DB 512 ATAAACAACCTTGGAGAGAGAAATC 534

RESULT 18  
ACN86065/c  
ID ACN86065 standard; DNA; 865 BP.  
XX AC ACN86065;  
XX 02-DEC-2004 (first entry)  
DT  
XX Breast cancer related marker, seq id 72115.  
DE  
XX Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.  
KW  
XX Homo sapiens.  
OS  
XX US2003099974-A1.  
PN  
XX 29-MAY-2003.  
PD  
XX 18-JUL-2002; 2002US-00198846.  
PF  
XX 18-JUL-2001; 2001US-0306220P.  
PR  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX Lillie J, Xu Y, Wang Y, Steinmann K;  
XX WPI; 2003-787014/74.  
DR  
XX Novel isolated polypeptide associated with breast cancer, useful for  
PT detecting presence of polypeptide in sample, as a marker for breast  
PT cancer.  
XX

XX Disclosure; SEQ ID NO 7215; 36pp; English.  
PS  
XX The invention relates to an isolated polypeptide (I) associated with a  
XX breast cancer which is encoded by a nucleic acid molecule comprising a  
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to  
CC the polypeptide of the invention. The activity of the polypeptide of the  
CC invention may be described as cytostatic. The antibody is useful for  
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the  
CC invention are useful in the detection of breast tumours. (I) is useful as  
CC a marker for breast cancer and in breast cancer therapy. Sequences given  
CC in records ACN78851-ACN92934 represent nucleic acid markers associated  
CC with breast cancer. Note: The sequence listing does not form part of the  
CC specification but may be obtained in electronic format from the USPTO web  
CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974  
XX  
SQ Sequence 865 BP; 227 A; 177 C; 174 G; 246 T; 0 U; 41 Other;

Query Match 72.8%; Score 18.2; DB 11; Length 865;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACAACCTTGGAGGGAGACT 23  
||| ||||| ||||| ||||| |||||  
DB 575 AAAAAGAAGCTTGGAGGTAAGT 553

RESULT 19  
ADSS57832  
ID ADSS57832 standard; cDNA; 1260 BP.  
XX AC  
XX ADS57832;  
XX  
XX DT 02-DEC-2004 (first entry)  
XX DE Bacterial polynucleotide #9819.  
XX  
XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.  
XX  
XX Bacteria.  
XX  
XX US2003233675-A1.  
XX  
XX PD 18-DEC-2003.  
XX  
XX PF 20-FEB-2003; 2003US-00369493.  
XX  
XX PR 21-FEB-2002; 2002US-0360039P.  
XX  
XX PA (CAOV/) CAO Y.  
XX PA (HINK/) HINKLE G J.  
XX PA (SLAT/) SLATER S C.  
XX PA (CHEN/) CHEN X.  
XX PA (GOLD/) GOLDMAN B S.  
XX  
XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
XX DR WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 33506; 122pp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
XX Sequence 1260 BP; 320 A; 281 C; 373 G; 286 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 13; Length 1260;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGAGGAGACT 23  
DB 25 AAACAACAGTTGGAGGAGGAGACT 47  
RESULT 20  
ABL68555  
ID ABL68555 standard; DNA; 2092 BP.  
XX  
XX ABL68555;  
XX  
XX DT 15-MAY-2002 (first entry)  
XX DE Kidney cancer related gene sequence SEQ ID NO:6892.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200194629-A2.  
XX  
XX PD 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US010838.  
XX  
XX 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 29-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.

```
PR 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 6892; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 9% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
XX Sequence 2092 BP; 728 A; 359 C; 443 G; 562 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 18.2; DB 6; Length 2092;
Best Local Similarity 87.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACT 23
Db 724 AAACACACTTGGCAAGGAGACT 746

RESULT 21
AD131610
ID AD131610 standard; cDNA; 2092 BP.
XX
XX AD131610;
XX
XX 17-JUN-2004 (first entry)
XX
XX Human cDNA #936.
XX
XX Human; gene; ss; immunological response; immunopathological condition;
KW Crohn's disease; asthma; ulcerative colitis; hyperesoinophilia;
KW irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;
KW acute monocytic leukaemia; antinflammatory; antiasthmatic; antiulcer;
KW osteopathic; antiarthritic; antirheumatic; cytostatic.
XX
XX Homo sapiens.
XX
XX US6607879-B1.
XX
XX 19-AUG-2003.
XX
XX 09-FEB-1998; 98US-00023655.
XX
XX 09-FEB-1998; 98US-00023655.
XX (INCY-) INCYTE CORP.
XX
XX Cocks BG, Stuart SG, Seilhamer JJ;
PI
```

```
XX WPI; 2003-895307/82.
XX
XX A composition comprising a plurality of cDNAs, useful for detecting
PT altered expression of genes in an immunological response or for
PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma
PT or osteoarthritis.
XX
XX Claim 1; SEQ ID NO 936; 50pp; English.
XX
XX The invention relates to a composition comprising a plurality of cDNAs
CC for detecting the altered expression of genes in an immunological
CC response. The invention also relates to a method of diagnosing or
CC monitoring the treatment of an immunopathological condition in a sample,
CC comprising obtaining nucleic acids from a sample, contacting the nucleic
CC acids of the sample with an array comprising the plurality of cDNAs under
CC conditions to form one or more hybridisation complexes, detecting the
CC hybridisation complexes and comparing the levels of the detected
CC hybridisation complexes with the level of hybridisation complexes
CC detected in a non-diseased sample, where an altered level of the detected
CC hybridisation complexes correlates with the presence of an
CC immunopathological condition. Also disclosed are an expression profile
CC comprising a microarray and a plurality of detectable complexes and a
CC method for identifying a plurality of polynucleotide probes. The cDNAs
CC are useful as hybridisable array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray can be used in
CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,
CC ulcerative colitis, hyperesoinophilia, irritable bowel syndrome,
CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in
CC identifying agents for the treatment of the diseases. The microarray may
CC also be used in drug discovery and development, toxicological and
CC carcinogenicity studies, forensics or pharmacogenomics. The composition
CC may also be used in purification of a subpopulation of mRNAs, cDNAs or
CC genomic fragments. This sequence represents a human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2092 BP; 728 A; 359 C; 443 G; 562 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 18.2; DB 11; Length 2092;
Best Local Similarity 87.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACT 23
Db 724 AAACACACTTGGCAAGGAGACT 746

RESULT 22
AD583677
ID AD583677 standard; cDNA; 2092 BP.
XX
XX AD583677;
XX
XX 11-AUG-2005 (first entry)
XX
XX Human lymph node cDNA #936.
XX
XX ss; gene; human; immunological response; blood cell; cancer;
KW immunopathological; AIDS; allergy; anaemia; asthma; atherosclerosis;
KW bronchitis; ulcerative colitis; diabetes; multiple sclerosis;
KW osteoporosis; pancreatitis; infection; arthritis; lymph node.
XX
XX Homo sapiens.
XX
XX US2004077003-A1.
XX
XX 22-APR-2004.
XX
XX 14-AUG-2003; 2003US-00641643.
XX
XX 09-FEB-1998; 98US-00023655.
XX
```



XX PA (INCY-) INCYTE CORP.  
 XX PI Cocks BG, Stuart SG, Seilhamer JJ;  
 XX DR WPI; 2004-387937/36.  
 XX PT New compositions having a number of first, second and third  
 XX PT polynucleotide probes, useful in research and diagnostic applications in  
 XX PT cancer and immunological conditions e.g. AIDS, diabetes, osteoporosis and  
 XX PT infections.  
 XX PS Claim 15; SEQ ID NO 936; 16pp; English.  
 XX CC The invention relates to polynucleotides which are used as probes to  
 CC detect genes differentially expressed in an immunological response,  
 CC abundantly expressed in an immunological response and/or coding for a  
 CC polypeptide known to regulate blood cell biology. The polynucleotides are  
 CC useful in research and diagnostic applications particularly in cancer and  
 CC immunopathological conditions, such as AIDS, allergies, anaemia, asthma,  
 CC atherosclerosis, bronchitis, ulcerative colitis, diabetes, multiple  
 CC sclerosis, osteoporosis, pancreatitis, infections and arthritis. The  
 CC present sequence represents a human lymph node cDNA used to detect blood  
 CC cell and immunological response gene expression. Note: The present  
 CC sequence does not appear in the printed specification but was obtained in  
 CC electronic format from the USPTO web site  
 CC (seqdata.uspto.gov/sequence.html?DocID=20040077003).  
 XX SQ Sequence 2092 BP; 728 A; 359 C; 443 G; 562 T; 0 U; 0 Other;  
 Query Match 72.8%; Score 18.2; DB 13; Length 2092;  
 Best Local Similarity 87.0%; Pred. No. 3.3e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAACACTTGGGAAGGAGACT 23  
 |||||  
 Db 724 AAACACACTTGGCAAGGAGACT 746  
 |||||  
 RESULT 23  
 ADT47914/c  
 ID ADT47914 standard; cDNA; 2306 BP.  
 XX AC ADT47914;  
 XX DT 02-DEC-2004 (first entry)  
 XX DE Bacterial polynucleotide #22665.  
 XX KW Recombinant DNA construct; transformed plant; improved plant property;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
 KW pathogen tolerance; pest tolerance; plant disease resistance;  
 KW cell cycle pathway modification; plant growth regulator;  
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KW bacterial polynucleotide; gene; ss.  
 XX OS Bacteria.  
 XX PN US2003233675-A1.  
 XX PD 18-DEC-2003.  
 XX PF 20-FEB-2003; 2003US-00369493.  
 XX PR 21-FEB-2002; 2002US-0360039P.  
 XX PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 XX WPI; 2004-061375/06.  
 XX PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX PS Claim 1; SEQ ID NO 46352; 122pp; English.  
 XX CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.  
 XX SQ Sequence 2306 BP; 689 A; 450 C; 447 G; 720 T; 0 U; 0 Other;  
 Query Match 72.8%; Score 18.2; DB 13; Length 2306;  
 Best Local Similarity 87.0%; Pred. No. 3.4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 AAAAAACACTTGGGAAGGAGACTC 24  
 |||||  
 Db 1161 ACAACACTAGGAAGGAGGCTC 1139  
 |||||  
 RESULT 24  
 ACA57281  
 ID ACA57281 standard; cDNA; 2307 BP.  
 XX AC ACA57281;  
 XX DT 10-JUN-2003 (first entry)  
 XX DE Human adipocyte Selected Interacting domain, SID, cDNA #368.  
 XX KW Human; ss; gene; prey; adipocyte; SID; selected interacting domain;  
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;  
 KW yeast 2-hybrid assay; metabolic disorder; obesity.  
 XX OS Homo sapiens.  
 XX PN WO200286122-A2.  
 XX PD 31-OCT-2002.  
 XX PF 14-MAR-2002; 2002WO-EP003768.  
 XX PR 14-MAR-2001; 2001US-0275734P.  
 XX PA (HYBR-) HYBRIGENICS.  
 XX PI Legrain P, Daviet L;

XX WPI; 2003-103412/09.  
DR P-PSDB; ABU70737.  
XX  
XX  
PT New complex between two interacting proteins in adipocyte cells, useful  
PT for identifying selected interacting domains that modulate protein  
PT interactions, or for preventing or treating metabolic disorders such as  
PT obesity or diabetes.  
XX  
XX Claim 7; Page 230-231; 382pp; English.  
PS  
XX The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence encodes a SID  
CC (prey) protein of the invention  
XX  
XX Sequence 2307 BP; 810 A; 392 C; 518 G; 587 T; 0 U; 0 Other;  
SQ

Query Match 72.8%; Score 18.2; DB 8; Length 2307;  
Best Local Similarity 87.0%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACT 23  
||| ||||| ||||| ||||| |||||  
Db 1297 AACAACACTTGGCAAGGAGACT 1319

RESULT 25  
AD031133  
ID ADO31133 standard; DNA; 6721 BP.  
XX AC ADO31133;  
XX  
XX 12-AUG-2004 (first entry)  
DT  
XX N-end rule ubiquitylation detection method gene #7.  
DE  
XX ds; gene; muscular; neuroprotective; nootropic; antiparkinsonian;  
KW cytosolic; gene therapy; ubiquitin; protein ubiquitylation; ataxia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW Frontotemporal Dementia; cancer; lymphoma; astrocyte differentiation;  
KW CNS injury.  
XX  
XX Homo sapiens.  
OS  
XX WO2004042352-A2.  
PN  
XX 21-MAY-2004.  
PD  
XX 28-OCT-2003; 2003WO-US034148.  
PF  
XX 30-OCT-2002; 2002US-0422448P.  
PR  
XX 12-JUL-2003; 2003US-0486529P.  
XX

PA (MESO-) MESO SCALE TECHNOLOGIES LLC.  
XX Davydov L, Kenten JH, Biebuyck H, Oberoi P;  
PI  
XX WPI; 2004-419747/39.  
DR P-PSDB; ADO31121.  
XX  
XX Isolated in vitro complex useful for treating ataxia, Alzheimer's  
PT disease, Parkinson's disease or cancer, comprises ubiquitin and a  
PT protein.  
PT  
XX Disclosure; SEQ ID NO 19; 137pp; English.  
PS  
XX The invention relates to an isolated in vitro complex comprising a  
CC ubiquitin or its derivative and a protein, e.g. aprataxin, tau, SLP,  
CC HMGI7, Pin1, Ctr, Cullin 3, HMG3, HSPC144, CDC6 or their fragments or  
CC derivatives, where the complex is formed via N-end rule ubiquitylation  
CC and the complex has a specific activity that is five-fold greater than  
CC the specific activity of in cell lysates, cells or tissue. The complex,  
CC methods, compositions, kits and compounds are useful in detecting,  
CC measuring and modulating protein ubiquitylation. The complex and the  
CC pharmaceutical compositions are useful in treating ataxia,  
CC neurodegenerative disorder, e.g. Alzheimer's disease, Parkinson's disease  
CC or Frontotemporal Dementia, cancer, human disorders and lymphomas and in  
CC improving astrocyte differentiation in response to CNS injury. This  
CC sequence corresponds to a gene used in the invention.  
XX  
XX Sequence 6721 BP; 2119 A; 1160 C; 1359 G; 2083 T; 0 U; 0 Other;  
SQ

Query Match 72.8%; Score 18.2; DB 12; Length 6721;  
Best Local Similarity 87.0%; Pred. No. 4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGACT 23  
||| ||||| ||||| ||||| |||||  
Db 1618 AACAACACTTGGCAAGGAGACT 1640

RESULT 26  
ADX97505  
ID ADX97505 standard; DNA; 6721 BP.  
XX AC ADX97505;  
XX  
XX 21-APR-2005 (first entry)  
DT  
XX Pancreatic cancer associated human protein encoding DNA, SEQ ID 53.  
DE  
XX pancreas tumor; cytostatic; gene; ds.  
KW  
XX Homo sapiens.  
OS  
XX EP1471075-A2.  
PN  
XX 27-OCT-2004.  
PD  
XX 31-MAR-2004; 2004EP-00090124.  
PF  
XX 31-MAR-2003; 2003DE-01015834.  
PR  
XX (HINZ/) HINZMANN B.  
PA (ROSE/) ROSENTHAL A.  
PA (PILA/) PILARSKY C.  
PA (DAHL/) DAHL E.  
PA (SPEC/) SPECHT T.  
PA (LICH/) LICHTNER R.  
XX  
XX Rosenthal A, Pilarsky C, Dahl E, Specht T, Bruemendorf T;  
PI Lichtner R, Staub E, Roepcke S, Li X;  
PI  
XX WPI; 2004-768082/76.  
DR P-PSDB; ADX97576.  
XX

PT New nucleic acid differentially expressed in pancreatic tumor tissue, for  
XX use as diagnostic agents and in screening for therapeutic agents.

PS Claim 1; SEQ ID NO 53; 28pp; German.

XX The invention relates to a novel human nucleic acid sequence of the  
CC pancreas and its encoded protein. The invention further comprises:  
CC proteins and peptides, preferably isolated, that contain a sequence  
CC encoded by the novel nucleic acid; and methods for diagnosis and  
CC treatment of pancreatic cancer, using a substance that inhibits or binds  
CC to the protein or its DNA, including: an antisense oligonucleotide, short  
CC interfering RNA or ribozyme directed against the pancreatic protein, an  
CC organic molecule, particularly having a molecular weight below 5000,  
CC especially 300, that binds to the pancreatic DNA, an aptamer or  
CC (monoclonal) antibody, preferably human or humanized, that binds to the  
CC pancreatic DNA, or an anti-idiotypic antibody raised against the  
CC monoclonal antibody, any of which may be derivatized with a reporter  
CC group, cytotoxic compound, immunostimulant and/or radioisotope. The novel  
CC human pancreatic proteins and their encoding DNA have cytostatic  
CC activity. The novel sequences are useful for inhibiting transcription  
CC and/or expression of genes and proteins associated with pancreatic  
CC cancer. This polynucleotide sequence represents the DNA encoding one of  
CC the novel human pancreatic proteins of the invention. Note: This sequence  
CC is not shown in the specification, it has been electronically downloaded  
CC from a DVD-ROM provided with this specification by the European Patent  
CC Office.

XX SQ Sequence 6721 BP; 2119 A; 1160 C; 1359 G; 2083 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 13; Length 6721;  
Best Local Similarity 87.0%; Pred. No. 4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGAGACT 23

Db 1618 AAACACACTTGGCAAGGAGACT 1640

RESULT 27

ADZ49257  
ID ADZ49257 standard; DNA; 6787 BP.

XX AC ADZ49257;

XX 30-JUN-2005 (first entry)

XX Insulin signaling pathway related gene, SEQ ID 586.

XX Antidiabetic; Gene Therapy; Non-insulin dependent diabetes;  
XX insulin resistance; Gene; ds.

XX Homo sapiens.

XX US2005085436-A1.

XX 21-APR-2005.

XX 08-JUL-2004; 2004US-00887553.

XX 08-JUL-2003; 2003US-0485883P.

XX (LIHH//) LI H.

XX (MAJJ//) MA J.

XX Li H, Ma J;

XX WPI; 2005-305194/31.

XX Treating, preventing or ameliorating pathological conditions associated  
PT with dysregulation of the insulin signaling pathway (ISP) comprises  
PT administering to a subject an amount of a modulator of any of the  
PT proteins regulated by ISP.

XX

PS Disclosure; SEQ ID NO 586; 70pp; English.

XX The present invention relates to a method for treating, preventing or  
CC ameliorating pathological conditions associated with dysregulation of the  
CC insulin signaling pathway (ISP). The method comprises administering to a  
CC subject a modulator for ISP-regulated proteins or a pharmaceutical  
CC composition comprising the described modulator. The method is useful for  
CC treating, preventing or ameliorating pathological conditions associated  
CC with dysregulation of the ISP such as Type II diabetes or Type A syndrome  
CC of insulin resistance. The present sequence is a human homolog of a  
CC Drosophila gene regulated by ISP. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20050085436.

XX SQ Sequence 6787 BP; 2134 A; 1172 C; 1388 G; 2093 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 14; Length 6787;  
Best Local Similarity 87.0%; Pred. No. 4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGAGACT 23

Db 1685 AAACACACTTGGCAAGGAGACT 1707

RESULT 28

ADQ97164  
ID ADQ97164 standard; DNA; 145985 BP.

XX AC ADQ97164;

XX 07-OCT-2004 (first entry)

XX Human cancer associated sequence HD1-08-009, SEQ ID 140.

XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.

XX Homo sapiens.

XX WO2004060304-A2.

XX 22-JUL-2004.

XX 22-DEC-2003; 2003WO-US041389.

XX 27-DEC-2002; 2002US-00330773.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-543781/52.

XX New isolated cancer associated nucleic acids comprising at least 10  
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
PT cancers such as leukemia and lymphoma.

XX Claim 1; SEQ ID NO 140; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-  
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 145985 BP; 40325 A; 30740 C; 31191 G; 43391 T; 0 U; 338 Other;

Query Match 72.8%; Score 18.2; DB 12; Length 145985;  
Best Local Similarity 87.0%; Pred. No. 6.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCTTGGAGGAGACT 23  
 ||||| ||||| ||||| ||||| |||||  
 Db 112891 AAAAAAAGCTTGGAGGAGACT 112913

RESULT 29  
 ADD17061  
 ID ADD17061 standard; DNA; 346 BP.  
 XX  
 AC ADD17061;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE DNA (SeqID 1129) that confers an altered visual phenotype in plants.  
 XX  
 KW ds; visual phenotype; plant; architecture; leaf surface; chlorotic;  
 KW bleaching; etching; wet leaf; stunting; elongation; texture;  
 KW agronomic trait; growth regulation; dwarf variety; insect resistance;  
 KW heat stress; transgenic.  
 XX  
 OS Unidentified.  
 XX  
 PN WO2003020741-A1.  
 XX  
 PD 13-MAR-2003.  
 XX  
 PF 30-AUG-2002; 2002WO-US027880.  
 XX  
 PR 31-AUG-2001; 2001US-0316326P.  
 XX  
 PA (DOWC ) DOW CHEM CO.  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX  
 PI Crosley R, Skokut T, Ruegger M, Larrinua I, Shukla V;  
 XX  
 DR WPI; 2003-300858/29.  
 XX  
 PT Novel isolated nucleic acid derived from Nicotiana benthamiana, Oryzae  
 PT sativa, Saccharomyces cerevisiae, and Papaver rhoeas, useful for  
 PT conferring altered visual phenotypes in plants.  
 XX  
 PS Claim 1; SEQ ID NO 1129; 517pp; English.  
 XX  
 CC This invention relates to the identification and isolation of novel  
 CC nucleic acid molecules that confer altered visual phenotypes in plants.  
 CC Specifically, it refers to modifications of plant architecture and/ or  
 CC leaf surface features in plants, such as chlorotic, bleaching, etching,  
 CC wet leaf, stunting, elongation and texture phenotypes, which are thought  
 CC will be agronomic traits beneficial to the farmer. As such, these novel  
 CC phenotypes can affect growth regulation i.e. useful for creating dwarf  
 CC varieties, exhibit resistance to insects or heat stress, confer changes  
 CC in pigment content such that plants have enhanced vitamin production or  
 CC delayed senescence and also for example produce plants that control the  
 CC production of ethylene. Furthermore, the present invention comprises  
 CC generating transgenic plants, as well as reproducibly altering the visual  
 CC phenotype of plant seeds, plant tissues and plant cells containing the  
 CC polynucleotides described herein. This polynucleotide is a homologue of a  
 CC DNA sequence that confers an altered visual phenotype when expressed in  
 CC plants, the method of the invention.  
 XX  
 SQ Sequence 346 BP; 111 A; 57 C; 84 G; 94 T; 0 U; 0 Other;  
 Query Match 71.2%; Score 17.8; DB 10; Length 346;  
 Best Local Similarity 90.5%; Pred. No. 3.9e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACACTTGGAGGAGACTCA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 105 AACACTTGGAGGAGACTCA 125

RESULT 30  
 ADC76449

ID ADC76449 standard; DNA; 524 BP.  
 XX  
 AC ADC76449;  
 XX  
 DT 01-JAN-2004 (first entry)  
 XX  
 DE DNA homologous to phytopathogen resistance-related cDNA - SEQ ID 1718.  
 XX  
 KW rice; yeast; poppy; plant; disease resistance; anti-fungal;  
 KW phytopathogen; gene shuffling; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN WO2003020905-A2.  
 XX  
 PD 13-MAR-2003.  
 XX  
 PF 30-AUG-2002; 2002WO-US027883.  
 XX  
 PR 31-AUG-2001; 2001US-0316392P.  
 XX  
 PA (DOWC ) DOW CHEM CO.  
 XX  
 PI Shukla V, Butler H, Larrinua I, Reddy AS;  
 XX  
 DR WPI; 2003-290185/28.  
 XX  
 PT Novel isolated nucleic acid derived from Nicotiana benthamiana, Oryzae  
 PT sativa, Saccharomyces cerevisiae, Trichoderma harzianum and Papaver  
 PT rhoeas, useful for conferring disease resistance in plants.  
 XX  
 PS Claim 1; SEQ ID NO 1718; 617pp; English.  
 XX  
 CC The invention relates to a novel isolated nucleic acid derived from  
 CC Nicotiana benthamiana, Oryza sativa (rice), Saccharomyces cerevisiae  
 CC (yeast), Trichoderma harzianum (Hypocrea lixii) and Papaver rhoeas  
 CC (poppy) and a sequence that hybridizes to them under conditions of low  
 CC stringency, where expression of the nucleic acid in a plant results in a  
 CC disease resistance phenotype. The polynucleotides of the invention  
 CC demonstrate anti-fungal activity and may be useful in conferring disease  
 CC resistance in a plant against phytopathogen such as Aspergillus flavus,  
 CC Gibberella fujikuroi and Gibberella zeae. Furthermore, the  
 CC polynucleotides may be useful to retrieve unknown sequences and in gene  
 CC shuffling or sexual PCR procedures. The current sequence is that of the  
 CC DNA of the invention which is homologous to that of the phytopathogen  
 CC resistance-related contig cDNAs.  
 XX  
 SQ Sequence 524 BP; 169 A; 69 C; 138 G; 148 T; 0 U; 0 Other;  
 Query Match 71.2%; Score 17.8; DB 10; Length 524;  
 Best Local Similarity 90.5%; Pred. No. 4.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACACTTGGAGGAGACTCA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 283 AACACTTGGAGGAGACTCA 303

RESULT 31  
 ADD17782  
 ID ADD17782 standard; DNA; 524 BP.  
 XX  
 AC ADD17782;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE DNA (SeqID 1850) that confers an altered visual phenotype in plants.  
 XX  
 KW ds; visual phenotype; plant; architecture; leaf surface; chlorotic;  
 KW bleaching; etching; wet leaf; stunting; elongation; texture;  
 KW agronomic trait; growth regulation; dwarf variety; insect resistance;  
 KW heat stress; transgenic.  
 XX



DR WPI; 2003-290185/28.  
XX Novel isolated nucleic acid derived from *Nicotiana benthamiana*, *Oryzae*  
PT *sativa*, *Saccharomyces cerevisiae*, *Trichoderma harzianum* and *Papaver*  
PT *rheas*, useful for conferring disease resistance in plants.  
XX  
PS Claim 1; SEQ ID NO 774; 617pp; English.  
XX  
CC The invention relates to a novel isolated nucleic acid derived from  
CC *Nicotiana benthamiana*, *Oryza sativa* (rice), *Saccharomyces cerevisiae*  
CC (yeast), *Trichoderma harzianum* (*Hypocrea lixii*) and *Papaver rhoeas*  
CC (poppy) and a sequence that hybridises to them under conditions of low  
CC stringency, where expression of the nucleic acid in a plant results in a  
CC disease resistance phenotype. The polynucleotides of the invention  
CC demonstrate anti-fungal activity and may be useful in conferring disease  
CC resistance in a plant against phytopathogens such as *Aspergillus flavus*,  
CC *Gibberella fujikuroi* and *Gibberella zeae*. Furthermore, the  
CC polynucleotides may be useful to retrieve unknown sequences and in gene  
CC shuffling or sexual PCR procedures. The current sequence is that of the  
CC DNA of the invention which is homologous to that of the phytopathogen  
CC resistance-related contig cDNAs.  
XX  
SQ Sequence 530 BP; 173 A; 69 C; 140 G; 148 T; 0 U; 0 Other;  
  
Query Match 71.2%; Score 17.8; DB 10; Length 530;  
Best Local Similarity 90.5%; Pred. No. 4.2e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 AACACTTGGAGGGAGACTCA 25  
DB 289 AACACTTGGATGAGACTCA 309  
|||||||  
  
RESULT 34  
ADB75375/c  
ID ADB75375 standard; cDNA; 5347 BP.  
XX  
AC ADB75375;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Prostate cancer marker cDNA.  
XX  
KW Prostate; cancer; cytostatic; gene therapy; marker; ss.  
XX  
OS *Homo sapiens*.  
XX  
PN WO2003009814-A2.  
XX  
PD 06-FEB-2003.  
XX  
PF 25-JUL-2002; 2002WO-US023913.  
XX  
PR 25-JUL-2001; 2001US-0307982P.  
PR 22-AUG-2001; 2001US-0314356P.  
PR 25-SEP-2001; 2001US-0325020P.  
PR 12-DEC-2001; 2001US-0341746P.  
PR 05-MAR-2002; 2002US-0362158P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;  
PI Hoerash S, Kamatkar S, Wonsley AM, Glatt K, Zhao X, Anderson D;  
XX  
XX WPI; 2003-248033/24.  
XX  
XX New nucleic acid molecule, useful for diagnosing or treating prostate  
PT cancer.  
XX  
PS Disclosure; SEQ ID NO 199; 99pp; English.  
XX  
CC The invention relates to newly discovered cancer markers associated with  
CC the cancerous state of prostate cells. Also disclosed is a method of

CC assessing whether a patient is afflicted with prostate cancer. The method  
CC of the invention involves assessing whether a patient is afflicted with  
CC prostate cancer by comparing the level of expression of a marker in a  
CC patient sample and the normal level of expression of the marker in a  
CC control non-prostate cancer sample, where a significant increase in the  
CC level of expression of the marker in the patient sample and the normal  
CC level indicates that the patient is afflicted with prostate cancer.  
CC Nucleic acids of the invention are useful for diagnosing or treating  
CC prostate cancer, and may be useful in gene therapy. Sequences given in  
CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5347 BP; 1268 A; 1439 C; 1368 G; 1272 T; 0 U; 0 Other;  
  
Query Match 71.2%; Score 17.8; DB 10; Length 5347;  
Best Local Similarity 90.5%; Pred. No. 5.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAAACAACACTTGGAGGGGAGA 21  
DB 1038 AAGAAACACATGGAAGGGAGA 1018  
|||||||  
  
RESULT 35  
ABD33003/c  
ID ABD33003 standard; DNA; 30642 BP.  
XX  
AC ABD33003;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Mouse cancer-associated genomic DNA MD21-026.  
XX  
KW Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;  
KW Leukaemia; lymphoma; CAP.  
XX  
OS *Mus musculus*.  
XX  
PN WO2004074320-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 17-FEB-2004; 2004WO-US004730.  
XX  
PR 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-00388838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
PA (SAGE-) SAGES DISCOVERY INC.  
XX  
PI Morris DW, Morris DW, Malandro MS;  
XX  
XX WPI; 2004-652914/63.  
XX  
PT New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX  
PS disclosure; seqid 773; 310pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-

CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 30642 BP; 7617 A; 6784 C; 6669 G; 9572 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 13; Length 30642;  
Best Local Similarity 90.5%; Pred. No. 7.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACACTTGGAGGGAGACTCA 25  
Db 21362 AACACTTGGAGGGAGAGACA 21342

RESULT 36  
ADQ97067/C  
ID ADQ97067 standard; DNA; 72352 BP.

XX AC ADQ97067;  
XX AC ADQ97067;  
XX 07-OCT-2004 (first entry)  
XX Mouse cancer associated sequence MD11-040, SEQ ID 43.  
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.  
XX Mus musculus.  
XX WO2004060304-A2.  
XX 22-JUL-2004.  
XX 22-DEC-2003; 2003WO-US0411389.  
XX 27-DEC-2002; 2002US-00330773.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
XX WPI; 2004-543781/52.  
XX New isolated cancer associated nucleic acids comprising at least 10  
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX cancers such as leukemia and lymphoma.  
XX Claim 1; SEQ ID NO 43; 199pp; English.  
XX The present invention relates to cancer associated sequences (ADQ97025-  
XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence

CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 72352 BP; 19658 A; 15474 C; 15337 G; 19982 T; 0 U; 1901 Other;  
Query Match 71.2%; Score 17.8; DB 12; Length 72352;  
Best Local Similarity 90.5%; Pred. No. 8.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGA 21  
Db 10509 AAAAGACACTTGGAGGGAGA 10489

RESULT 37  
AAD36511  
ID AAD36511 standard; DNA; 169998 BP.

XX AAD36511;  
XX 09-AUG-2002 (first entry)  
XX Human Her-1 gene.  
XX Human; epidermal growth factor receptor; hyperproliferative disease;  
XX Her1; psoriasis; psoriasis; tumour; cancer; gene; ds.

XX Homo sapiens.  
XX Key Location/Qualifiers  
FH 1208..1472  
FT exon /tag= a

FT intron 1473..124390  
FT exon /tag= b  
FT exon 124391..124544  
FT intron /tag= c  
FT intron 124545..125409  
FT exon /tag= d  
FT exon 125410..125595  
FT intron /tag= e  
FT intron 125596..128711  
FT exon /tag= f  
FT exon 128712..128848  
FT intron /tag= g  
FT intron 128849..133400  
FT exon /tag= h  
FT exon 133401..133469  
FT intron /tag= i  
FT exon 133470..134652  
FT exon /tag= j  
FT exon 134653..134773  
FT intron /tag= k  
FT exon 134774..136116  
FT exon /tag= l  
FT intron /tag= m  
FT exon 136117..136261  
FT intron /tag= n  
FT exon 136262..137936  
FT exon /tag= o  
FT intron 137937..138053  
FT exon /tag= p  
FT exon 138054..138637  
FT exon /tag= q  
FT intron 138638..138766  
FT exon /tag= r  
FT intron 138767..138864  
FT exon /tag= s  
FT intron 138865..138940  
FT exon /tag= t  
FT intron 138941..139765  
FT exon /tag= u



FT intron 139861. .142245  
FT /tag= v  
FT exon 142246. .142445  
FT /tag= w  
FT intron 142446. .143605  
FT /tag= x  
FT exon 143606. .143738  
FT /tag= y  
FT intron 143739. .145838  
FT /tag= z  
FT exon 145839. .145931  
FT /tag= aa  
FT intron 145932. .147385  
FT /tag= ab  
FT exon 147386. .147544  
FT /tag= ac  
FT intron 147545. .153274  
FT /tag= ad  
FT exon 153275. .153321  
FT /tag= ae  
FT intron 153322. .155088  
FT /tag= af  
FT exon 155089. .155231  
FT /tag= ag  
FT intron 155232. .156025  
FT /tag= ah  
FT exon 156026. .156151  
FT /tag= ai  
FT intron 156152. .156826  
FT /tag= aj  
FT exon 156827. .156928  
FT /tag= ak  
FT intron 156929. .163399  
FT /tag= al  
FT exon 163400. .163586  
FT /tag= am

WO200226758-A1.

04-APR-2002.

28-SEP-2001; 2001WO-US030551.

29-SEP-2000; 2000US-00676610.

(ISIS-) ISIS PHARM INC.

Bennett CF, Wyatt JR, Freier SM;

WPI; 2002-394234/42.

XX Novel antisense oligonucleotide that specifically hybridizes with and  
XX inhibits nucleic acid encoding epidermal growth factor receptor, useful  
XX for treating hyperproliferative disease such as cancer or psoriasis.

PS Example 19; Page 67-121; 169pp; English.

XX The invention relates to an antisense oligonucleotide targetted to a  
CC nucleic acid molecule encoding human epidermal growth factor receptor  
CC (Her1) to inhibit its expression. The antisense compounds are useful for  
CC treating diseases or conditions associated with Her-1 such as  
CC hyperproliferative diseases especially cancer (lung, ovarian, colon or  
CC prostate cancer) and psoriasis. They are also useful as research  
CC reagents, diagnostics, therapeutics, kits and prophylactically e.g. to  
CC prevent or delay tumour formation. The present sequence is human Her-1  
CC gene

XX Sequence 169998 BP; 46143 A; 38164 C; 37751 G; 47940 T; 0 U; 0 Other;

Query Match

Best Local Similarity 71.2%; Score 17.8; DB 6; Length 169998;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGAC 22  
Db 104105 AAAAGACTTGGAGGGAGAC 104125  
||||| ||||| ||||| |||||

RESULT 38

ABN85584

ID ABN85584 standard; DNA; 197496 BP.

XX AC ABN85584;

XX DT 09-SEP-2002 (first entry)

XX Human EGFR SEQ ID NO 10.

XX Human; EGFR; HER2-neu; chemotherapeutic regimen; tumour; cancer;  
KW receptor tyrosine kinase; epidermal growth factor receptor;  
KW gene expression; ds.

XX OS Homo sapiens.

XX PN WO200244413-A2.

XX PD 06-JUN-2002.

XX PF 09-NOV-2001; 2001WO-US043035.

XX PR 01-DEC-2000; 2000US-0250122P.

XX PR 04-DEC-2000; 2000US-0250469P.

XX PR 11-JUN-2001; 2001US-00877177.

XX PA (RESP-) RESPONSE GENETICS INC.

XX PI Danenberg KD;

XX DR WPI; 2002-537460/57.

XX Determining chemotherapeutic regimen of receptor tyrosine kinase targeted  
XX agent for treating tumor by examining EGFR and/or HER2-neu mRNA amount in  
XX tumor cells, comparing it to predetermined threshold expression level.

PS Disclosure; Page 71-124; 125pp; English.

XX The invention relates to determining the chemotherapeutic regimen of  
CC receptor tyrosine kinase targeted agent for treating tumour by amplifying  
CC mRNA from tumour and non-malignant tissues using a primer pair that  
CC hybridises to epidermal growth factor receptor (EGFR) and/or HER2-neu  
CC gene (i), quantitating and obtaining differential expression levels of  
CC amplified mRNA and comparing the differential expression levels and  
CC threshold levels for expression of (i). The method is useful for  
CC assessment of clinical treatment of a patient and as a diagnostic or  
CC prognostic tool for a range of cancers including breast, head and neck,  
CC lung, oesophageal and colorectal cancer. The present sequence is that of  
CC the human EGFR DNA sequence used in methods of the invention

XX Sequence 197496 BP; 53640 A; 44528 C; 43228 G; 56100 T; 0 U; 0 Other;

Query Match

Best Local Similarity 71.2%; Score 17.8; DB 6; Length 197496;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGAC 22

Db 112105 AAAAGACTTGGAGGGAGAC 112125  
||||| ||||| ||||| |||||

RESULT 39

ABD32854

ID ABD32854 standard; DNA; 209083 BP.

XX AC ABD32854;

XX DT 18-NOV-2004 (first entry)

XX DE Human cancer-associated genomic DNA HD17-032.  
XX KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;  
XX KW leukaemia; lymphoma; CAP.  
XX OS Homo sapiens.  
XX PN WO2004074320-A2.  
XX PD 02-SEP-2004.  
XX PF 17-FEB-2004; 2004WO-US004730.  
XX PR 14-FEB-2003; 2003US-00367094.  
XX PR 14-MAR-2003; 2003US-00388838.  
XX PR 15-APR-2003; 2003US-00417375.  
XX PR 13-JUN-2003; 2003US-00461862.  
XX PR 15-SEP-2003; 2003US-00663431.  
XX PR 15-SEP-2003; 2003US-00737318.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX PA Morris DW, Morris DW, Malandro MS;  
XX PI WPI; 2004-652914/63.  
XX DR New isolated cancer-associated polynucleotides and polypeptides useful  
XX FT for diagnosing, preventing or treating cancers, especially lymphoma and  
XX FT leukemia, or in screening for agents that modulate cancer.  
XX FT  
XX PS claim 16; seqid 524; 310pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising at least 10  
XX CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
XX CC in the specification, or its complement. The nucleic acids encode cancer-  
XX CC associated proteins. Also included are an expression vector comprising  
XX CC the isolated nucleic acid cited above, a host cell comprising the above  
XX CC recombinant nucleic acid or expression vector, a microarray for detecting  
XX CC a cancer-associated (CA) nucleic acid comprising at least one probe  
XX CC comprising at least 10 contiguous nucleotides of any of the above-  
XX CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
XX CC an open reading frame of a CA sequence selected from any of the 95  
XX CC polynucleotide sequences as mentioned in the specification, or its  
XX CC complement), an isolated antibody, (or its antigen binding fragment) that  
XX CC binds to the above polypeptide, a hybridoma that produces the above  
XX CC monoclonal antibody, a pharmaceutical composition comprising the above  
XX CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
XX CC cells (comprising the antibody cited above, methods for diagnosing cancer  
XX CC or for detecting the presence or absence of cancer cells in an  
XX CC individual, a method for inhibiting growth of cancer cells in an  
XX CC individual, a method for delivering a therapeutic agent to cancer cells  
XX CC in an individual, an electronic library comprising the above  
XX CC polynucleotide or polypeptide (or their fragments), methods of screening  
XX CC for anticancer activity or for a bioactive agent capable of modulating  
XX CC the activity of a CA protein (CAP), methods for detecting cancer  
XX CC associated with expression of a polypeptide in a test cell sample, a  
XX CC method for treating cancers and a method for inhibiting the expression of  
XX CC CA gene in a cell. The composition and methods are useful for detecting,  
XX CC diagnosing, preventing and treating cancers, especially lymphoma and  
XX CC leukaemia. These may also be used in screening for agents that modulate  
XX CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
XX CC sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 209083 BP; 57124 A; 47057 C; 45679 G; 59223 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 13; Length 209083;  
Best Local Similarity 90.5%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAGACTTGGAGGGAGAC 22

Db 112983 AAAAAGACTTGGAGGGAGAC 113003  
RESULT 40  
ADC24763/c  
ID ADC24763 standard; DNA; 334462 BP.  
XX AC ADC24763;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human wild type HNL4X genomic sequence.  
XX KW ds; nootropic; neuroleptic; tranquilizer; gene therapy; synaptogenesis;  
XX KW mutation; neurological disease; mental disorder; psychiatric illness;  
XX KW autism; Asperger syndrome; schizophrenia;  
XX KW attention deficit hyperactivity disorder.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT exon 10670..10828  
FT /tag= a  
FT /number= 1b  
FT intron 10829..83743  
FT /tag= b  
FT /number= 1b  
FT exon 83744..84513  
FT /tag= c  
FT /number= 2  
FT intron 84514..201675  
FT /tag= d  
FT /number= 2  
FT exon 201676..201828  
FT /tag= e  
FT /number= 3  
FT intron 201829..240358  
FT /tag= f  
FT /number= 3  
FT exon 240359..240469  
FT /tag= g  
FT /number= 3b  
FT intron 240470..312939  
FT /tag= h  
FT /number= 3b  
FT exon 312940..313125  
FT /tag= i  
FT /number= 4  
FT intron 313126..318312  
FT /tag= j  
FT /number= 4  
FT exon 318313..319102  
FT /tag= k  
FT /number= 5  
FT intron 319103..327017  
FT /tag= l  
FT /number= 5  
FT exon 327018..330406  
FT /tag= m  
FT /number= 6  
XX PN WO2003045998-A2.  
XX PD 05-JUN-2003.  
XX PF 02-DEC-2002; 2002WO-FR004134.  
XX PR 30-NOV-2001; 2001CA-02364106.  
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX PA (INSP ) INST PASTEUR.  
XX PA (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.

```
XX Bourgeron T, Jamaïn S, Quach H, Betancur C, Leboyer M;
PI Gillberg C;
XX WPI; 2003-493399/46.
DR New nucleic acid encoding mutant protein involved in synaptogenesis,
PT useful for treatment and diagnosis of e.g. autism, Asperger syndrome, and
PT schizophrenia.
XX Claim 10; SEQ ID NO 1; 416pp; French.
XX The invention relates to an isolated or purified polynucleotide encoding
XX a polypeptide (the wild-type form of which is involved in synaptogenesis)
XX that includes at least one mutation associated with development of
XX neurological disease and/or a predisposition to development of mental
XX disorders or psychiatric illness. The polypeptide are used to screen for
XX agents that modulate their activity. Also nucleic acid, polypeptide,
XX polypeptide-specific antibodies, vectors containing he nucleic acid and
XX host cells containing the vector, are useful as pharmaceuticals for
XX treating mental and neurological disorders, specifically autism, Asperger
XX syndrome, schizophrenia and attention deficit hyperactivity disorder. The
XX wild-type forms of the nucleic acid and polypeptide can be used
XX similarly. Also detecting mutations in the nucleic acid and polypeptide,
XX or measuring activity of the polypeptide, can be used to detect
XX biochemical disorders that affect formation of synapses and to diagnose
XX mental disease. This sequence corresponds to the genomic sequence of the
XX human wild type HNL4X gene.
XX Sequence 334462 BP; 97253 A; 64494 C; 66298 G; 106417 T; 0 U; 0 Other;
SQ Query Match 71.2%; Score 17.8; DB 10; Length 334462;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAACACTTGGAGGGAGA 21
DB 198779 AATACACTTGGAGGGAGA 198759
RESULT 41
AAC13345/C ;
ID AAC13345 standard; cDNA; 120 BP.
XX AAC13345;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 17420.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
```

```
PS Claim 1; SEQ ID NO 17420; 71pp + Sequence Listing; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX Sequence 120 BP; 29 A; 25 C; 23 G; 43 T; 0 U; 0 Other;
SQ Query Match 70.4%; Score 17.6; DB 3; Length 120;
Best Local Similarity 83.3%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACACTTGGAGGGAGACTC 24
DB 114 AAAAAAACTTGTAAAGTTAGACTC 91
RESULT 42
AAV04596
ID AAV04596 standard; DNA; 299 BP.
XX AAV04596;
XX 25-MAR-2003 (revised)
DT 02-JUL-1998 (first entry)
XX Flea serine protease DNA sequence SEQ ID NO:127.
XX Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
KW immunoglobulin protease; larvae; host animal; ss.
XX Siphonaptera.
XX WO9740058-A1.
XX 30-OCT-1997.
XX 24-APR-1997; 97WO-US006121.
XX 24-APR-1996; 96US-00639075.
PR 15-NOV-1996; 96US-00749699.
PR 04-APR-1997; 97US-0042945P.
XX (HESK-) HESKA CORP.
XX Grieve RB, Rushlow KE, Hunter SW, Frank GR, Steigler GL;
PI Gaines PJ, Silver G;
XX WPI; 1998-076762/07.
XX New flea protease genes and proteins - used in vaccine compositions for
XX the prophylaxis and treatment of flea infestation, especially in cats or
XX dogs.
XX Claim 1; Page 259; 318pp; English.
XX The present sequence is a DNA sequence from a novel flea serine protease.
XX The protease, its mimetopes, antibodies (Ab) and inhibitors of the
XX protein, as well as the DNA encoding the protein, may all be used in
XX therapeutic compositions to reduce flea protease activity (especially
XX immunoglobulin protease) and so reduce flea infestation, especially in
XX cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
XX
```

CC fleas which comprises anti-protease antibodies produced by a host animal  
 CC in response to administration of the protein. Therapeutic compositions  
 CC may further comprise a compound that reduces haematophagous ectoparasite  
 CC burden by a method other than by reducing flea immunoglobulin protease  
 CC activity. The novel flea DNA encoding the protein can also be used to  
 CC produce recombinant protein, and fragments of it are used as probes and  
 CC primers for identification and isolation of related sequences, also as  
 CC antisense, triplex-forming agents and ribozymes for inhibition of the  
 CC synthesis of the protein. Ab are also useful for screening expression  
 CC libraries, to purify the protein and to target cytotoxins to fleas.  
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX  
 SQ Sequence 299 BP; 94 A; 45 C; 79 G; 79 T; 0 U; 2 Other;

Query Match 70.4%; Score 17.6; DB 2; Length 299;  
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACACTTGGGAAGGGAGACTCA 25  
 |||||  
 Db 267 AAAGACACTTGCAGGGAGATTCA 290

## RESULT 43

AAAC90858  
 ID AAC90858 standard; cDNA; 299 BP.

XX AAC90858;

XX 19-MAR-2001 (first entry)

XX Flea serine protease nfsP8-299 nucleotide sequence #92.

XX Flea; cat flea; serine protease; aminopeptidase; cysteine protease;  
 KW flea infestation; proteolytic; insecticide; vaccine; cat; dog; dog flea;  
 KW Ctenocephalides felis; Ctenocephalides canis; Pulex irritans;  
 KW ectoparasite; ss.

XX Ctenocephalides felis.

XX US6150125-A.

XX 21-NOV-2000.

XX 24-APR-1996; 96US-00639075.

XX 13-DEC-1991; 91US-00806482.

XX 18-OCT-1994; 94US-00326773.

XX 07-JUN-1995; 95US-00482130.

XX 07-JUN-1995; 95US-00484211.

XX 07-JUN-1995; 95US-00485443.

XX 15-AUG-1997; 97WO-US014442.

XX (HESK-) HESKA CORP.

XX Gaines PJ, Silver G, Rushlow KE, Hunter SW, Frank GR;

PI Stiegler GL, Grieve RB;

XX WPI; 2001-136374/14.

XX P-PSDB; AAB50602.

XX New isolated flea proteins with proteolytic activity, useful for  
 PT preventing and reducing flea infestations in mammals especially cats and  
 PT dogs.

XX Claim 4; Col 153-154; 150pp; English.

XX The present invention describes isolated flea serine protease,  
 CC aminopeptidase and cysteine protease proteins (I). Also described is a  
 CC method for identifying a compound (II) capable of inhibiting flea  
 CC protease activity comprising: (a) contacting (I) with a protease  
 CC substrate and a putative inhibitory compound, where (I) has proteolytic

CC activity in the absence of the compound; and (b) determining if the  
 CC compound inhibits protease activity by detecting cleavage of the protease  
 CC substrates; where decreased cleavage of the protease substrate indicates  
 CC an inhibitory compound. (I), nucleic acid molecules encoding (I), and  
 CC antibodies immunospecific for (I) and (II) are useful for preventing and  
 CC reducing flea infestations, particularly the species Ctenocephalides  
 CC felis (cat fleas), Ctenocephalides canis (dog fleas) and Pulex irritans,  
 CC in animals, preferably cats and dogs. They are also useful for reducing  
 CC infestation by other ectoparasites, preferably mosquitoes, midges,  
 CC sandflies, blackflies, ticks and Rhodnius. AAC90818 to AAC90913 and  
 CC AAB50551 to AAB50644 represent sequences used in the exemplification of  
 CC the present invention

XX  
 SQ Sequence 299 BP; 94 A; 45 C; 79 G; 79 T; 0 U; 2 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 299;  
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACACTTGGGAAGGGAGACTCA 25  
 |||||  
 Db 267 AAAGACACTTGCAGGGAGATTCA 290

## RESULT 44

AAA00406/c

ID AAA00406 standard; cDNA; 300 BP.

XX AAA00406;

XX 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:397.

XX Human; colon cancer; tumour; diagnosis; gene expression product; probe;  
 KW detection; cancerous state; metastasis; identification; breast cancer;  
 KW oestrogen receptor-positive breast cancer; therapy;  
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO9558675-A2.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US010602.

XX 14-MAY-1998; 98US-0085426P.

XX 15-MAY-1998; 98US-0085537P.

XX 15-MAY-1998; 98US-0085696P.

XX 21-OCT-1998; 98US-0105234P.

XX 27-OCT-1998; 98US-0105877P.

XX (CHIR ) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leashkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX WPI; 2000-126369/11.

XX Polynucleotide library used to determine cancerous states of mammalian  
 PT cells.

XX Claim 1; Page 275; 1097pp; English.

XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
 CC libraries constructed from human colon cancer cell lines. The present  
 CC invention also describes a method of detecting differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell, comprising  
 CC detecting at least one differentially expressed gene product in a test

CC sample derived from a cell suspected of being cancerous, where detection  
CC of the differentially expressed gene product is correlated with a  
CC cancerous state of the cell from which the test sample was derived. The  
CC polynucleotide sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of pre-  
CC metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer  
XX  
SQ Sequence 300 BP; 88 A; 37 C; 47 G; 128 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 3; Length 300;  
Best Local Similarity 83.3%; Pred. NO. 4.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 AAAAAACACTTGGAGGGAGACTC 24  
Db 294 AGAAAAACACTTGGAGGAAAGCTC 271  
  
RESULT 45  
AAF65920/c  
ID AAF65920 standard; cDNA; 389 BP.  
XX  
AC AAF65920;  
XX  
XX  
DT 09-APR-2001 (first entry)  
XX  
DE Novel human polynucleotide, SEQ ID NO: 1676.  
XX  
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
KW breast cancer; lung cancer; cancer detection; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200102568-A2.  
PN  
PD 11-JAN-2001.  
XX  
XX 30-JUN-2000; 2000WO-US018374.  
PF  
XX 02-JUL-1999; 99US-0142310P.  
PR  
XX 02-JUL-1999; 99US-0142311P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;  
PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;  
PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;  
XX  
XX WPI; 2001-091805/10.  
DR  
XX  
XX Library of polynucleotides for diagnosing a cancerous state of a  
PT mammalian cell and detecting cancer, particularly of the colon or  
PT prostate, comprises 3351 human polynucleotide sequences.  
PT  
XX  
PS Claim 9; Page 784; 1046pp; English.  
XX  
XX The present sequence is one of 3351 sequences in a library of human  
CC polynucleotides. The library is used to detect differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell and can  
CC detect colon, prostate, breast and lung cancer. The library can be used  
CC to produce probes for detection of mRNA and to produce additional copies  
CC of the polynucleotides. The probes can be used for chromosome mapping of  
CC the polynucleotide and for detection of transcription levels. Ribozymes  
CC or antisense oligonucleotides can be generated. The polynucleotides and  
CC their gene products are used as genetic or biochemical markers (e.g. in

CC blood or tissues) that will detect the earliest changes along the  
CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
CC preventive interventions. The polynucleotides, polypeptides and  
CC antibodies against them can be used in pharmaceutical compositions to  
CC treat the cancers and proliferative disorders such as neoplasia,  
CC dysplasia and hyperplasia  
XX  
SQ Sequence 389 BP; 109 A; 87 C; 93 G; 99 T; 0 U; 1 Other;  
  
Query Match 70.4%; Score 17.6; DB 5; Length 389;  
Best Local Similarity 83.3%; Pred. NO. 4.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 AAAAAACACTTGGAGGGAGACTC 24  
Db 242 ATAAAGGACTTGGAGAGAGACTC 219  
  
RESULT 46  
AAI88689/c  
ID AAI88689 standard; cDNA; 398 BP.  
XX  
AC AAI88689;  
XX  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 8749.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200164835-A2.  
PN  
PD 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US004927.  
PF  
XX 28-FEB-2000; 2000US-00515126.  
PR  
XX 18-MAY-2000; 2000US-00577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-514838/56.  
DR  
XX P-PSDB; AAO08758.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.  
PT  
XX Claim 1; SEQ ID NO 8749; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from Wipo at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 398 BP; 123 A; 69 C; 74 G; 131 T; 0 U; 1 Other;  
XX  
XX Query Match 70.4%; Score 17.6; DB 4; Length 398;  
Query Match







Query Match 70.4%; Score 17.6; DB 14; Length 480;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGAAGGAGACTC 24  
| | | | | | | | | | | | | | | | | | | | | |  
Db 239 ATAAAGGACTTGAAGGAGAGACTC 216

Search completed: February 3, 2006, 21:56:41  
Job time : 211.111 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttggaaggagactca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_est9.\*

10: gb\_est10.\*

11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	450	4	AK007298 Mus muscu
2	23.4	93.6	460	5	BY707400 BY707400
3	20.8	83.2	298	5	BX631074 BX631074
4	20.8	83.2	464	2	BT796159 BT796159
5	20.8	83.2	466	2	BF454696 BF454696
6	20.8	83.2	623	10	CL346540 CL346540
7	20.8	83.2	666	2	BF454299 BF454299
8	20.8	83.2	669	2	BF538629 BF538629
9	20.8	83.2	680	2	BG872391 BG872391
10	20.8	83.2	682	2	BG865082 BG865082
11	20.8	83.2	683	2	BG865198 BG865198
12	20.8	83.2	737	2	BG866078 BG866078
13	20.8	83.2	747	2	BG866256 BG866256
14	20.8	83.2	755	2	BG866256 BG866256
15	20.8	83.2	764	7	CK294820 CK294820
16	20.8	83.2	775	2	BG870364 BG870364
17	20.8	83.2	856	2	BG865927 BG865927
18	20.8	83.2	881	2	BF302440 BF302440
19	20.8	83.2	904	2	BF538107 BF538107
20	20.8	83.2	928	2	BF537576 BF537576
21	20.8	83.2	952	2	BF302117 BF302117
22	20.2	80.8	711	9	BZ023877 BZ023877

96	18.8	75.2	932	7	CK402743	AUF IfInt
97	18.8	75.2	953	8	DN577957	92246236
98	18.8	75.2	983	9	CC331869	CGUEN69TV
99	18.8	75.2	1187	8	DN701279	CLJ34-G01
100	18.8	75.2	1334	10	AG442056	Mus muscu
101	18.8	75.2	2829	4	AK081536	Mus muscu
102	18.6	74.4	141	7	CV307827	t47b09.b
103	18.6	74.4	141	7	CV307828	t47b09.g
104	18.6	74.4	336	7	CK713568	zF201-P00
105	18.6	74.4	351	8	CK355204	ycag7d08..
106	18.6	74.4	400	3	BI790314	1c96h07.x
107	18.6	74.4	403	8	DR903915	JGI_XZT65
108	18.6	74.4	416	8	DR542611	WS01039..B
109	18.6	74.4	440	6	CD311383	StrPu621..
110	18.6	74.4	446	1	AA403014	zC63612.f
111	18.6	74.4	452	5	BU671855	HaT423 He
112	18.6	74.4	459	1	AA398499	zt63612..s
113	18.6	74.4	508	5	BQ386941	NISC mn21
114	18.6	74.4	512	9	AZ098943	RPCI-23-4
115	18.6	74.4	516	9	AQ998152	RPCI-23-2
116	18.6	74.4	517	9	BZ330770	hV97a08.g
117	18.6	74.4	517	9	AQ685663	HS_5556 A
118	18.6	74.4	536	5	BX116611	BX116611
119	18.6	74.4	543	3	BM257260	520815 MA
120	18.6	74.4	545	6	CD332039	StrPu537..
121	18.6	74.4	558	1	AW441225	EST310621
122	18.6	74.4	577	9	CE130490	tigr-g88-
123	18.6	74.4	584	9	AZ640544	IM0502C09
124	18.6	74.4	593	9	BH281646	CH230-49G
125	18.6	74.4	609	9	AZ080614	RPCI-23-4
126	18.6	74.4	618	10	CZ094521	OM_Ba009
127	18.6	74.4	619	10	CE525050	tigr-g88-
128	18.6	74.4	621	8	CV950784	PvrPv43
129	18.6	74.4	636	6	CD307379	StrPu691..
130	18.6	74.4	652	11	CR085510	Reverse s
131	18.6	74.4	654	10	CB413313	tigr-g88-
132	18.6	74.4	659	6	CB530282	737067 WA
133	18.6	74.4	663	10	CE537115	tigr-g88-
134	18.6	74.4	667	10	CE752213	tigr-g88-
135	18.6	74.4	673	7	CK950704	4089959 B
136	18.6	74.4	679	6	CD336366	StrPu537..
137	18.6	74.4	680	6	CD310869	StrPu691..
138	18.6	74.4	689	10	CE677158	tigr-g88-
139	18.6	74.4	709	9	CC390114	PHT0127B
140	18.6	74.4	718	9	CE138835	tigr-g88-
141	18.6	74.4	744	8	DR439912	EST149..18
142	18.6	74.4	752	10	AG376010	Mus muscu
143	18.6	74.4	763	1	AJ814183	AJ814183
144	18.6	74.4	776	2	BG203562	RST22946
145	18.6	74.4	780	10	AG603339	Mus muscu
146	18.6	74.4	783	2	BG213532	RST33139
147	18.6	74.4	784	10	CL581889	OB_Ba004
148	18.6	74.4	799	2	BG189152	RST8192 A
149	18.6	74.4	838	10	DU039116	22937 Tom
150	18.6	74.4	865	9	BZ722675	PUDAL237B

## ALIGNMENTS

RESULT 1  
AK007298 450 bp mRNA linear HTC 03-APR-2004  
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched  
DEFINITION library, clone:1700127D06 product:similar to tissue kallikrein [EC  
3.4.21.35], submandibular mCK-2 (fragment) [Mus musculus], full  
insert sequence.  
ACCESSION AK007298  
VERSION AK007298.1 GI:12840752  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861  
4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 450)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,  
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,  
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Nishi, K.,  
Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,  
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,  
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,  
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.jp/>) for further  
details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence [5',  
GAGAGAGAGCGCCGCAATTAATTCGAGTTAATTAATTAATTCCTCCCTCC 3']. cDNA  
was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'  
end: SstI. Host: SOLR.  
Location/Qualifiers  
FEATURES

[illegible]

Query Match 93.6%; Score 23.4; DB 5; Length 460;  
Best Local Similarity 96.0%; Pred. No. 12;  
Matches 24: Conservative 0; Mismatches 1; Indels

QY 1 AAAAAACACTTGGAAAGGAGACTCA 25

```

Db      230 AAAAAAAGACTTGGAGGGAGACTCA 254
RESULT 3
BX631074/c
LOCUS   BX631074.1 GI:33610946
DEFINITION
Mus musculus (house mouse)
ACCESSION
BX631074
VERSION
BX631074.1
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 298)
Hentrich,J., Hermans,J., Kranz,H., Loebbert,R., Schluter,T.,
Schuette,D., Weindel,M., Heil,O., Ebert,L., Neubert,P., Peters,M.,
Radlof,U., Schneider,D. and Korn,B.
Mouse ArrayTAG cDNA (LION)
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD: LIONp462B0412.
RZPDLib;
Mouse ArrayTAG cDNA (LION)
http://www.rzpd.de/cgi-bin/products/showlib.pl.cgi/response?libNo=4
62 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
Contact RZPD (clone@rzpd.de) for further information. Seq primer:
RP: CAGGAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..298
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="LIONp462B0412"
/lab_host="DH10B"
/clone_lib="pBluescript Lion"
ORIGIN
Query Match 83.2%; Score 20.8; DB 5; Length 298;
Best Local Similarity 91.7%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAACACTTGGAGGGAGACTCA 25
|||||
Db 218 AAAGACACTTGCNAGGGAGACTCA 195
|||||
RESULT 4
BB796159
LOCUS   BB796159
DEFINITION
BB796159 RIKEN full-length enriched, Leydig cells CRL-2065 MLTC-1
cDNA Mus musculus cDNA clone G4D0006D17 3', mRNA sequence.
ACCESSION
BB796159
VERSION
BB796159.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 464)
Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Mateuyma,T., Nakamura,M., Nishi,K., Nomura,K., Numaaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akchira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Mateuyma,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1..464
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G4D0006D17"
/tissue_type="testis"
/cell_type="Leydig cells"
/cell_lines="CRL-2065 MLTC-1"
/clone_lib="RIKEN full-length enriched, Leydig cells
CRL-2065 MLTC-1 cDNA"
ORIGIN
Query Match 83.2%; Score 20.8; DB 2; Length 464;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAACACTTGGAGGGAGACTCA 25
|||||
Db 254 AAAGACACTTGTAGGGAGACTCA 277
|||||
RESULT 5
BF454696
LOCUS   BF454696
DEFINITION
BF454696
similar to SW:KJXL_MOUSE P15948 GLANDULAR KALLIKREIN K22 PRECURSOR
; mRNA sequence.
ACCESSION
BF454696
VERSION
BF454696.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Euthera; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
 1 (bases 1 to 466)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@mail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

MGI:1452521  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 418.

FEATURES  
 source Location/Qualifiers

1. .466  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3810409"  
 /sex="male"  
 /tissue\_type="salivary gland"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI CGAP SGI"  
 /notes="vector: PCMV-SPORT6; Site 1: NotI; Site 2: SalI;  
 Cloned unidirectionally. Primer: Oligo dT. Average insert  
 size 1.6 Kb. Library constructed by Life Technologies."

## ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 466;  
 Best Local Similarity 91.7%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACTCA 25

DB 314 AAAGACACTTGTAGGGAGACTCA 337

RESULT 6  
 CL346540  
 LOCUS CL346540 623 bp DNA linear GSS 19-AUG-2004  
 DEFINITION RPC144\_270F2\_r RPC1-44 Sus scrofa genomic clone RPC144\_270F2,  
 genomic survey sequence.

ACCESSION CL346540  
 VERSION CL346540.1 GI:51398509  
 KEYWORDS GSS.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
 Sus.

REFERENCE 1 (bases 1 to 623)  
 Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,  
 Beaver, J.B., and Schook, L.B.

TITLE Piggy-BACing the Human Genome: Constructing a Porcine Physical Map  
 Through Comparative Genomics

JOURNAL Unpublished (2004)  
 COMMENT Other GSSs: RPC144\_270F2.f

Contact: Lawrence B. Schook  
 Department of Animal Sciences  
 University of Illinois at Urbana Champaign  
 1201 W. Gregory Dr., Urbana, IL 61801, USA  
 Tel: 217 265 5326  
 Fax: 217 244 5617  
 Email: schook@uiuc.edu

Clones are derived from the porcine BAC library RPC1-44

(<http://www.bacpac.chori.org/porcine242.htm>). For BAC library  
 availability, please contact Pieter de Jong (pdejong@chori.org).  
 Clones may be purchased from BACPAC Resources  
 (<http://BACPACorders.chori.org>). This work was undertaken as part  
 of the International Swine Genome Sequencing Consortium by  
 University of Illinois at Urbana Champaign, USA with funds provided  
 by grant No. AG2002-34480-11828 from USDA-CSREES and  
 AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing  
 Initiative)  
 Plate: 270 row: F column: 2  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source

Location/Qualifiers  
 1. .623  
 /organism="Sus scrofa"  
 /mol\_type="genomic DNA"  
 /strain="four pigs (breed: 37.5% Yorks Landrace and 25%  
 Meishan)"  
 /db\_xref="taxon:9823"  
 /clone="RPC144\_270F2"  
 /sex="male"  
 /cell\_type="blood"  
 /clone\_lib="RPC1-44"  
 /notes="vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;  
 porcine male BAC library produced by Pieter de Jong"

## ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 623;  
 Best Local Similarity 91.7%; Pred. No. 1.9e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACTCA 25

DB 258 AAAACACATGGAAGGGAGACTCA 281

## RESULT 7

BF454299

LOCUS BF454299 666 bp mRNA linear EST 01-DEC-2000

DEFINITION maa03c06.y1 NCI CGAP\_SGI Mus musculus cDNA clone IMAGE:3810034 5',  
 similar to SW:KIK9\_MOUSE P15949 GLANDULAR KALLIKREIN K9 PRECURSOR  
 ;, mRNA sequence.

ACCESSION BF454299

VERSION BF454299.1 GI:11520468

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 666)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other ESTs: maa03c06.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

MGI:1452146

Seq primer: -40RP from Gibco

High quality sequence stop: 439.

## FEATURES

source

Location/Qualifiers  
 1. .666  
 /organism="Mus musculus"  
 /mol\_type="mRNA"



```

/db_xref="taxon:10090"
/clone="IMAGE:3810034"
/sex="male"
/tissue_type="salivary gland"
/dev_stage="5 months"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP SG1"
/notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.6 kb. Library constructed by Life Technologies."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 666;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25
Db 643 AAAGACACTTGCAGGGAGACTCA 666

RESULT 8
BF538629 602052947F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4192307 5',
LOCUS mRNA sequence.
DEFINITION
ACCESSION BF538629
VERSION
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 669)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9522 row: m column: 12
High quality sequence stop: 667.
Location/Qualifiers
1. 669
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="IMAGE:4192307"
/clone="IMAGE:4192307"
/clone_lib="NCI CGAP_SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 669;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25
Db 450 AAAGACACTTGTAAAGGAGACTCA 473

RESULT 9
BG872391 60279684F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4923716 5',
LOCUS mRNA sequence.
DEFINITION
ACCESSION BG872391
VERSION
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 680)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM10845 row: h column: 21
High quality sequence stop: 678.
Location/Qualifiers
1. 680
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="IMAGE:4923716"
/clone="IMAGE:4923716"
/clone_lib="NCI CGAP_SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 680;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25
Db 639 AAAGACACTTGTAAAGGAGACTCA 662

RESULT 10
BG865082 602784236F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4910264 5',
LOCUS mRNA sequence.
DEFINITION
ACCESSION BG865082
VERSION
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 682)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.

```

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10810 row: h column: 09  
 High quality sequence stop: 679.  
 Location/Qualifiers

1. .682  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4910264"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
 NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

## FEATURES

Source

## ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 682;  
 Best Local Similarity 91.7%; Pred. No. 2e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25

DB 461 AAAGACACTTGAAGGAGACTCA 484

## RESULT 11

BG869198

LOCUS

DEFINITION 6027899082F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4920164 5',  
 mRNA sequence.

ACCESSION BG869198

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 683)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10836 row: d column: 21

High quality sequence stop: 660.

Location/Qualifiers

1. .683

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4920164"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI\_CGAP\_SG2"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
 NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 683;  
 Best Local Similarity 91.7%; Pred. No. 2e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25

DB 452 AAAGACACTTGAAGGAGACTCA 475

## RESULT 12

BG865647

LOCUS

DEFINITION 602783810F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4909778 5',  
 mRNA sequence.

ACCESSION BG865647

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 737)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10809 row: d column: 03

High quality sequence stop: 736.

Location/Qualifiers

1. .737

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4909778"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI\_CGAP\_SG2"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
 NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
 dt. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 737;  
 Best Local Similarity 91.7%; Pred. No. 2e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGAGACTCA 25

DB 643 AAAGACACTTGAAGGAGACTCA 666

## RESULT 13

BG866078

LOCUS

DEFINITION 602787987F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4914236 5',  
 mRNA sequence.

ACCESSION BG866078

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





```

/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 881;
Best Local Similarity 91.7%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGGAGACTCA 25
   |||||
Db 634 AAAGACACTTGTGAAGGAGACTCA 657

RESULT 19
BF538107
LOCUS
DEFINITION
602053573F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4192434 5',
mRNA sequence.
ACCESSION
BF538107
VERSION
BF538107.1 GI:11625475
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 904)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9523 row: b column: 19
High quality sequence stop: 768.
Location/Qualifiers
1..904
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4192434"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 904;
Best Local Similarity 91.7%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGGAGACTCA 25
   |||||
Db 572 AAAGACACTTGTGAAGGAGACTCA 595

RESULT 20
BF537576
LOCUS
DEFINITION
602053339F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4192689 5',
mRNA sequence.
ACCESSION
BF537576
VERSION
BF537576.1 GI:11624944
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 928)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9523 row: m column: 10
High quality sequence stop: 522.
Location/Qualifiers
1..928
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4192689"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_SG2"
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match      83.2%; Score 20.8; DB 2; Length 928;
Best Local Similarity 91.7%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGGAGACTCA 25
   |||||
Db 265 AAAGACACTTGTGAAGGAGACTCA 288

RESULT 21
BF302117
LOCUS
DEFINITION
602033067F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4167953 5',
mRNA sequence.
ACCESSION
BF302117
VERSION
BF302117.1 GI:11248640
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 952)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

http://image.llnl.gov  
Plate: LLAM9459 row: f column: 18  
High quality sequence stop: 830.  
Location/Qualifiers

# FEATURES

source

1. .952  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4167953"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP SG2"  
/notes="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

# ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 952;  
Best Local Similarity 91.7%; Pred. No. 2.1e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGAAGGGAGACTCA 25  
|||||  
Db 616 AAAGACACTTGTAAAGGGAGACTCA 639

# RESULT 22

BZ023877/c

# LOCUS

DEFINITION BZ023877 711 bp DNA linear GSS 08-OCT-2002  
oej6901.b1 B.oleracea002 Brassica oleracea genomic, genomic survey  
sequence.

# ACCESSION

VERSION BZ023877

# KEYWORDS

# SOURCE

# ORGANISM

Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

# REFERENCE

AUTHORS Delehaenty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,  
Nash,W., Rabinowicz,P.D. and Wilson,R.K.

# TITLE

# JOURNAL

# COMMENT

Whole genome shotgun reads from Brassica oleracea  
Unpublished (2002)  
Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: submissions@wustl.edu  
Plate: oej69 row: g column: 01  
Seq primer: -21uppt forward  
Class: shotgun  
High quality sequence start: 16  
High quality sequence stop: 529.  
Location/Qualifiers

# FEATURES

source

1. .711  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3712"  
/clone\_lib="B.oleracea002"  
/note="Vector: pOTw13; Whole genome shotgun library from  
flowering buds. DNA was purified from a crude nuclear  
prep using Brassica oleracea T01000DH3 buds provided by  
Thomas Osborn at the University of Wisconsin. Genomic  
DNA was provided by Pablo Rabinowicz (CSHL) and the  
shotgun library prepared at Washington University Genome  
Sequencing Center."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 711;  
Best Local Similarity 88.0%; Pred. No. 3.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGACTCA 25  
|||||  
Db 122 AAAAAAATTTGAAGGGAGACTTA 98

# RESULT 23

BG866962

# LOCUS

DEFINITION BG866962 730 bp mRNA linear EST 29-MAY-2001  
602789830F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:492110 5',  
mRNA sequence.

# ACCESSION

VERSION BG869962

# KEYWORDS

# SOURCE

# ORGANISM

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

# REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

# TITLE

# JOURNAL

# COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10838 row: 1 column: 07  
High quality sequence stop: 730.  
Location/Qualifiers

# FEATURES

source

1. .730  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4921110"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP\_SG2"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:  
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 730;  
Best Local Similarity 88.0%; Pred. No. 3.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGACTCA 25  
|||||  
Db 634 AACAGACACTTGTAAAGGGAGACTCA 658

# RESULT 24

BG866966

# LOCUS

DEFINITION BG866966 719 bp mRNA linear EST 29-MAY-2001  
602785688F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4911981 5',  
mRNA sequence.

# ACCESSION

VERSION BG866966

# KEYWORDS

# SOURCE

# ORGANISM

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

# REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

# TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)





```

ACCESSION DN792152
VERSION DN792152.1 GI:62382219
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 878)
AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
TITLE A large-scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
JOURNAL Development 128 (13), 2615-2627 (2001)
PUBMED 11493577
COMMENT Contact: Erica Sodergren
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-7676
Fax: 713-798-6977
Email: ericas@bcm.tmc.edu
NCBI Trace Archive: 490875627
Insert Length: 1750 Std Error: 0.25
Plate: 131 row: K column: 22.
FEATURES             source
    Location/Qualifiers
        1..878
            /organism="Strongylocentrotus purpuratus"
            /mol_type="mRNA"
            /db_xref="taxon:7668"
            /clone="PMCSPP2-131K22"
            /tissue_type="embryo"
            /cell_type="primary mesenchyme cells"
            /lab_host="E.coli"
            /clone_lib="Sea Urchin primary mesenchyme cell cDNA
            library"
            /notes="Vector: pSPORT1; Site 1: NotI; Site 2: MluI; oligo
            dt priming from poly A+ RNA, directionally cloned"
ORIGIN
    1
    2
    3
    4
    5
    6
    7
    8
    9
    10
    11
    12
    13
    14
    15
    16
    17
    18
    19
    20
    21
    22
    23
    24
    25
    26
    27
    28
    29
    30
    31
    32
    33
    34
    35
    36
    37
    38
    39
    40
    41
    42
    43
    44
    45
    46
    47
    48
    49
    50
    51
    52
    53
    54
    55
    56
    57
    58
    59
    60
    61
    62
    63
    64
    65
    66
    67
    68
    69
    70
    71
    72
    73
    74
    75
    76
    77
    78
    79
    80
    81
    82
    83
    84
    85
    86
    87
    88
    89
    90
    91
    92
    93
    94
    95
    96
    97
    98
    99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
    701
    702
    703
    704
    705
    706
    707
    708
    709
    710
    711
    712
    713
    714
    715
    716
    717
    718
    719
    720
    721
    722
    723
    724
    725
    726
    727
    728
    729
    730
    731
    732
    733
    734
    735
    736
    737
    738
    739
    740
    741
    742
    743
    744
    745
    746
    747
    748
    749
    750
    751
    752
    753
    754
    755
    756
    757
    758
    759
    760
    761
    762
    763
    764
    765
    766
    767
    768
    769
    770
    771
    772
    773
    774
    775
    776
    777
    778
    779
    780
    781
    782
    783
    784
    785
    786
    787
    788
    789
    790
    791
    792
    793
    794
    795
    796
    797
    798
    799
    800
    801
    802
    803
    804
    805
    806
    807
    808
    809
    810
    811
    812
    813
    814
    815
    816
    817
    818
    819
    820
    821
    822
    823
    824
    825
    826
    827
    828
    829
    830
    831
    832
    833
    834
    835
    836
    837
    838
    839
    840
    841
    842
    843
    844
    845
    846
    847
    848
    849
    850
    851
    852
    853
    854
    855
    856
    857
    858
    859
    860
    861
    862
    863
    864
    865
    866
    867
    868
    869
    870
    871
    872
    873
    874
    875
    876
    877
    878
    879
    880
    881
    882
    883
    884
    885
    886
    887
    888
    889
    890
    891
    892
    893
    894
    895
    896
    897
    898
    899
    900
    901
    902
    903
    904
    905
    906
    907
    908
    909
    910
    911
    912
    913
    914
    915
    916
    917
    918
    919
    920
    921
    922
    923
    924
    925
    926
    927
    928
    929
    930
    931
    932
    933
    934
    935
    936
    937
    938
    939
    940
    941
    942
    943
    944
    945
    946
    947
    948
    949
    950
    951
    952
    953
    954
    955
    956
    957
    958
    959
    960
    961
    962
    963
    964
    965
    966
    967
    968
    969
    970
    971
    972
    973
    974
    975
    976
    977
    978
    979
    980
    981
    982
    983
    984
    985
    986
    987
    988
    989
    990
    991
    992
    993
    994
    995
    996
    997
    998
    999
    1000
    1001
    1002
    1003
    1004
    1005
    1006
    1007
    1008
    1009
    1010
    1011
    1012
    1013
    1014
    1015
    1016
    1017
    1018
    1019
    1020
    1021
    1022
    1023
    1024
    1025
    1026
    1027
    1028
    1029
    1030
    1031
    1032
    1033
    1034
    1035
    1036
    1037
    1038
    1039
    1040
    1041
    1042
    1043
    1044
    1045
    1046
    1047
    1048
    1049
    1050
    1051
    1052
    1053
    1054
    1055
    1056
    1057
    1058
    1059
    1060
    1061
    1062
    1063
    1064
    1065
    1066
    1067
    1068
    1069
    1070
    1071
    1072
    1073
    1074
    1075
    1076
    1077
    1078
    1079
    1080
    1081
    1082
    1083
    1084
    1085
    1086
    1087
    1088
    1089
    1090
    1091
    1092
    1093
    1094
    1095
    1096
    1097
    1098
    1099
    1100
    1101
    1102
    1103
    1104
    1105
    1106
    1107
    1108
    1109
    1110
    1111
    1112
    1113
    1114
    1115
    1116
    1117
    1118
    1119
    1120
    1121
    1122
    1123
    1124
    1125
    1126
    1127
    1128
    1129
    1130
    1131
    1132
    1133
    1134
    1135
    1136
    1137
    1138
    1139
    1140
    1141
    1142
    1143
    1144
    1145
    1146
    1147
    1148
    1149
    1150
    1151
    1152
    1153
    1154
    1155
    1156
    1157
    1158
    1159
    1160
    1161
    1162
    1163
    1164
    1165
    1166
    1167
    1168
    1169
    1170
    1171
    1172
    1173
    1174
    1175
    1176
    1177
    1178
    1179
    1180
    1181
    1182
    1183
    1184
    1185
    1186
    1187
    1188
    1189
    1190
    1191
    1192
    1193
    1194
    1195
    1196
    1197
    1198
    1199
    1200
    1201
    1202
    1203
    1204
    1205
    1206
    1207
    1208
    1209
    1210
    1211
    1212
    1213
    1214
    1215
    1216
    1217
    1218
    1219
    1220
    1221
    1222
    1223
    1224
    1225
    1226
    1227
    1228
    1229
    1230
    1231
    1232
    1233
    1234
    1235
    1236
    1237
    1238
    1239
    1240
    1241
    1242
    1243
    1244
    1245
    1246
    1247
    1248
    1249
    1250
    1251
    1252
    1253
    1254
    1255
    1256
    1257
    1258
    1259
    1260
    1261
    1262
    1263
    1264
    1265
    1266
    1267
    1268
    1269
    1270
    1271
    1272
    1273
    1274
    1275
    1276
    1277
    1278
    1279
    1280
    1281
    1282
    1283
    1284
    1285
    1286
    1287
    1288
    1289
    1290
    1291
    1292
    1293
    1294
    1295
    1296
    1297
    1298
    1299
    1300
    1301
    1302
    1303
    1304
    1305
    1306
    1307
    1308
    1309
    1310
    1311
    1312
    1313
    1314
    1315
    1316
    1317
    1318
    1319
    1320
    1321
    1322
    1323
    1324
    1325
    1326
    1327
    1328
    1329
    1330
    1331
    1332
    1333
    1334
    1335
    1336
    1337
    1338
    1339
    1340
    1341
    1342
    1343
    1344
    1345
    1346
    1347
    1348
    1349
    1350
    1351
    1352
    1353
    1354
    1355
    1356
    1357
    1358
    1359
    1360
    1361
    1362
    1363
    1364
    1365
    1366
    1367
    1368
    1369
    1370
    1371
    1372
    1373
    1374
    1375
    1376
    1377
    1378
    1379
    1380
    1381
    1382
    1383
    1384
    1385
    1386
    1387
    1388
    1389
    1390
    1391
    1392
    1393
    1394
    1395
    1396
    1397
    1398
    1399
    1400
    1401
    1402
    1403
    1404
    1405
    1406
    1407
    1408
    1409
    1410
    1411
    1412
    1413
    1414
    1415
    1416
    1417
    1418
    1419
    1420
    1421
    1422
    1423
    1424
    1425
    1426
    1427
    1428
    1429
    1430
    1431
    1432
    1433
    1434
    1435
    1436
    1437
    1438
    1439
    1440
    1441
    1442
    1443
    1444
    1445
    1446
    1447
    1448
    1449
    1450
    1451
    1452
    1453
    1454
    1455
    1456
    1457
    1458
    1459
    1460
    1461
    1462
    1463
    1464
    1465
    1466
    1467
    1468
    1469
    1470
    1471
    1472
    1473
    1474
    1475
    1476
    1477
    1478
    1479
    1480
    1481
    1482
    1483
    1484
    1485
    1486
    1487
    1488
    1489
    1490
    1491
    1492
    1493
    1494
    1495
    1496
    1497
    1498
    1499
    1500
    1501
    1502
    1503
    1504
    1505
    1506
    1507
    1508
    1509
    1510
    1511
    1512
    1513
    1514
    1515
    1516
    1517
    1518
    1519
    1520
    1521
    1522
    1523
    1524
    1525
    1526
    1527
    1528
    1529
    1530
    1531
    1532
    1533
    1534
    1535
    1536
    1537
    1538
    1539
    1540
    1541
    1542
    1543
    1544
    1545
    1546
    1547
    1548
    1549
    1550
    1551
    1552
    1553
    1554
    1555
    1556
    1557
    1558
    1559
    1560
    1561
    1562
    1563
    1564
    1565
    1566
    1567
    1568
    1569
    1570
    1571
    1572
    1573
    1574
    1575
    1576
    1577
    1578
    1579
    1580
    1581
    1582
    1583
    1584
    1585
    1586
    1587
    1588
    1589
    1590
    1591
    1592
    1593
    1594
    1595
    1596
    1597
    1598
    1599
    1600
    1601
    1602
    1603
    1604
    1605
    1606
    1607
    1608
    1609
    1610
    1611
    1612
    1613
    1614
    1615
    1616
    1617
    1618
    1619
    1620
    1621
    1622
    1623
    1624
    1625
    1626
    1627
    1628
    1629
    1630
    1631
    1632
    1633
    1634
    1635
    1636
    1637
    1638
    1639
    1640
    1641
    1642
    1643
    1644
    1645
    1646
    1647
    1648
    1649
    1650
    1651
    1652
    1653
    1654
    1655
    1656
    1657
    1658
    1659
    1660
    1661
    1662
    1663
    1664
    1665
    1666
    1667
    1668
    1669
    1670
    1671
    1672
    1673
    1674
    1675
    1676
    1677
    1678
    1679
    1680
    1681
    1682
    1683
    1684
    1685
    1686
    1687
    1688
    1689
    1690
    1691
    1692
    1693
    1694
    1695
    1696
    1697
    1698
    1699
    1700
    1701
    1702
    1703
    1704
    1705
    1706
    1707
    1708
    1709
    1710
    1711
    1712
    1713
    1714
    1715
    1716
    1717
    1718
    1719
    1720
    1721
    1722
    1723
    1724
    1725
    1726
    1727
    1728
    1729
    1730
    1731
    1732
    1733
    1734
    1735
    1736
    1737
    1738
    1739
    1740
    1741
    1742
    1743
    1744
    1745
    1746
    1747
    1748
    1749
    1750
    1751
    1752
    1753
    1754
    1755
    1756
    1757
    1758
    1759
    1760
    1761
    1762
    1763
    1764
    1765
    1766
    1767
    1768
    1769
    1770
    1771
    1772
    1773
    1774
    1775
    1776
    1777
    1778
    1779
    1780
    1781
    1782
    1783
    1784
    1785
    1786
    1787
    1788
    1789
    1790
    1791
    1792
    1793
    1794
    1795
    1796
    1797
    1798
    1799
    1800
    1801
    1802
    1803
    1804
    1805
    1806
    1807
    1808
    1809
    1810
    1811
    1812
    1813
    1814
    1815
    1816
    1817
    1818
    1819
    1820
    1821
    1822
    1823
    1824
    1825
    1826
    1827
    1828
    
```

Query Match 79.2%; Score 19.8; DB 8; Length 963;  
 Best Local Similarity 91.3%; Pred. No. 5.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACT 23  
 |||||  
 Db 209 AAAAAACATGGAGGAGACT 187

RESULT 30  
 DN661091  
 LOCUS  
 DEFINITION CEC50-E06.yld-s SHGC-CEC Gasterosteus aculeatus cDNA clone EST 29-MAR-2005  
 CEC50-E06 5', mRNA sequence.  
 DN661091  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Gasterosteus aculeatus (three spined stickleback)  
 Gasterosteus aculeatus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
 Gasterosteidae; Gasterosteus.

REFERENCE 1 (bases 1 to 1203)  
 Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,  
 Schmutz, J., and Myers, R.M.  
 Expressed sequence tags from Gasterosteus aculeatus  
 Unpublished (2003)  
 Contact: Grimwood, Jane  
 Stanford Human Genome Center  
 Stanford University School of Medicine  
 975 S California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 320 5917  
 Fax: 650 320 5801  
 Email: jane@shgc.stanford.edu  
 Plate: 50

High quality sequence stop: 711.  
 Location/Qualifiers  
 1. .1203  
 /organism="Gasterosteus aculeatus"  
 /mol\_type="mRNA"  
 /strain="Conner Creek sticklebacks, WA USA"  
 /db\_xref="taxon:69293"  
 /clone="CEC50-E06"  
 /sex="mixed male and female"  
 /tissue\_type="skin"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="SHGC-CEC"  
 /notes="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTCTAGACGAGCGGCCCTT25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at:  
<http://www.openbiosystems.com/cdna.library.constructionfaq.php#8>  
 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:  
<http://www.openbiosystems.com/stickleback>"

FEATURES  
 source  
 1. .1203  
 /organism="Gasterosteus aculeatus"  
 /mol\_type="mRNA"  
 /strain="Conner Creek sticklebacks, WA USA"  
 /db\_xref="taxon:69293"  
 /clone="CEC50-E06"  
 /sex="mixed male and female"  
 /tissue\_type="skin"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="SHGC-CEC"  
 /notes="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTCTAGACGAGCGGCCCTT25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at:  
<http://www.openbiosystems.com/cdna.library.constructionfaq.php#8>  
 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:  
<http://www.openbiosystems.com/stickleback>"

ORIGIN  
 Query Match 79.2%; Score 19.8; DB 8; Length 1203;

Best Local Similarity 91.3%; Pred. No. 6e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAAAAACACTTGGAGGAGACTC 24  
 |||||  
 Db 1125 AAAAAAACTTGGAGGAACTC 1147

RESULT 31  
 CW916421/c  
 LOCUS  
 DEFINITION CW916421 411 bp DNA linear GSS 08-DEC-2004  
 RPCI42\_160K4.TJ RPCI-42 Bos taurus genomic clone RPCI42\_160K4,  
 genomic survey sequence.  
 CW916421  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bos taurus (cow)  
 Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 Pecora; Bovidae; Bovinae; Bos.  
 1 (bases 1 to 411)  
 Larkin, D.M., Donthu, K., Lebuc, R., Ryan, K., Liu, L., de Jong, P.J. and  
 Lewin, H.A.

REFERENCE 1  
 End sequencing of Holstein BAC library RPCI-42  
 Unpublished (2004)  
 Other\_GSSs: RPCI42\_160K4.TV  
 Contact: Harris Lewin  
 Department of Animal Sciences  
 University of Illinois at Urbana Champaign  
 1201 W. Gregory Dr., Urbana, IL 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu  
 Clones are derived from the bovine BAC library RPCI-42  
 (http://bacpac.choi.org/mbovine42.htm). For BAC library  
 availability, please contact Pieter de Jong  
 (pdejong@mail.choi.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/ordering/information.htm).  
 Funding was provided in part by grant no. AG2004-34480-14417 from  
 USDA-CSREES (Livestock Genome Sequencing Initiative) and  
 AG58-5438-2-313 from USDA-ARS  
 Plate: 160 row: K column: 4  
 Seq primer: SP6  
 Class: BAC ends.  
 Location/Qualifiers  
 1. .411  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Holstein"  
 /db\_xref="taxon:9913"  
 /clone="RPCI42\_160K4"  
 /sex="male"  
 /cell\_type="blood"  
 /clone\_lib="RPCI-42"  
 /note="Vector: pBac3.6; Site 1: EcoRI; Site 2: EcoRI;  
 RPCI-42 Bovine BAC library (Mäle) produced by Pieter de  
 Jong"

FEATURES  
 source  
 1. .411  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Holstein"  
 /db\_xref="taxon:9913"  
 /clone="RPCI42\_160K4"  
 /sex="male"  
 /cell\_type="blood"  
 /clone\_lib="RPCI-42"  
 /note="Vector: pBac3.6; Site 1: EcoRI; Site 2: EcoRI;  
 RPCI-42 Bovine BAC library (Mäle) produced by Pieter de  
 Jong"

ORIGIN  
 Query Match 77.6%; Score 19.4; DB 10; Length 411;  
 Best Local Similarity 95.2%; Pred. No. 7.8e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 AACACTTGAAGGGAGACTCA 25  
 |||||  
 Db 282 AACACTTAGAGGGAGACTCA 262

RESULT 32  
 CNS02ES8  
 LOCUS  
 DEFINITION CNS02ES8 745 bp DNA linear GSS 01-SEP-2000  
 Tetraodon nigroviridis genome survey sequence T7 end of clone

131E09 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AL194129  
AL194129.1 GI:7832235  
GSS; genome survey sequence.  
Tetraodon nigroviridis

Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS

1  
Roest Crolius H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C.,  
Barnot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W., and Weissenbach, J.

TITLE  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence

JOURNAL  
PUBMED

Nat. Genet. 25 (2), 235-238 (2000)

REFERENCE  
AUTHORS

2  
Roest Crolius H., Jallou, O., Dasilva, C., Ozouf-Costaz, C.,  
Fizes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
Saurin, W., Barnot, A., and Weissenbach, J.

TITLE  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

JOURNAL  
PUBMED

Genome Res. 10 (7), 939-949 (2000)

REFERENCE  
AUTHORS

3 (bases 1 to 745)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.

FEATURES  
source

1..745  
Location/Qualifiers  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="131E09"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG131AC051P1  
end : 77"

ORIGIN

Query Match 77.6%; Score 19.4; DB 10; Length 745;  
Best Local Similarity 95.2%; Pred. No. 8.6e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAACAACCTTGGGAAGGGAGA 21  
Db 238 AAAAACAACCTTGGGAAGGGAGA 258

RESULT 33  
LOCUS

BI283322/c  
DEFINITION  
UI-R-DA0-bzf-g-06-0-UI.s1 UI-R-DA0 Rattus norvegicus CDNA clone  
UI-R-DA0-bzf-g-06-0-UI 3', mRNA sequence.

ACCESSION  
VERSION

BI283322  
BI283322.1 GI:14934954

KEYWORDS  
SOURCE

EST.  
Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 303)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene

JOURNAL  
PUBMED  
COMMENT

discovery  
Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized rat eye library cDNA library preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

FEATURES  
source

Location/Qualifiers

1..303  
/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-DA0-bzf-g-06-0-UI"

/dev\_stage="ADULT"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-DA0"

/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DA0  
library is a non-normalized library constructed from rat  
salivary gland tissue. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
been previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)

TAG TISSUE=rat eye

TAG\_LIB=UI-R-DA0

TAG\_SEQ=CAGCC"

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 303;  
Best Local Similarity 87.5%; Pred. No. 9.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACAACCTTGGGAAGGGAGACTCA 25

Db 234 AAAGACACTTGCACGAGGTGACTCA 211

RESULT 34

BU096127

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BU096127 306 bp mRNA linear EST 14-MAR-2003  
tca-612 tca Trypanosoma carassii CDNA clone 03m9 5', mRNA sequence.

BU096127

BU096127.1 GI:25123851

EST.

Trypanosoma carassii

Trypanosoma carassii

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 306)

Agarwal, F., Campo, V., Cremona, L., Jager, A., Di Noia, J.M.,

Overath, P., Sanchez, D.O. and Frasch, A.C.

Gene discovery in the freshwater fish parasite Trypanosoma

carassii: identification of trans-sialidase-like and mucin-like

genes

Infect. Immun. 70 (12), 7140-7144 (2002)

Contact: Sanchez DO

Genomics and Bioinformatics

Instituto de Investigaciones Bioteconologicas

Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos Aires, Argentina  
Tel: (54-11) 4580/7255/7  
Fax: (54-11) 4752-9639  
Email: dsanchez@ib.unsam.edu.ar  
Sequences were basecalled with phred and vector was masked with crossmatch (see <http://www.phrap.org>). Sequences were then trimmed from both ends to remove low quality bases and masked vector.  
Plate: 03 row: m column: 9  
Seq primer: R7.

FEATURES  
source

Location/Qualifiers  
1..306  
/organism="Trypanosoma carassii"  
/mol\_type="mrna"  
/db\_xref="taxon:38249"  
/clone="03m9"  
/dev\_stage="blood trypanostigote"  
/lab\_host="Goldfish (Carassius auratus)"  
/clone\_lib="tca"  
/note="Vector: pSport1; Blood trypanostigotes were obtained from goldfish and cultured as described (Overath et al. Parasitol Res (1998) 84:343) before obtaining total RNA using Trizol. cDNA library construction was made from polyA+ mRNA using a poly-dT oligonucleotide as primer. The cDNAs were cloned in a oriented manner using a commercial kit (SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning, Life Technologies)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 5; Length 306;  
Best Local Similarity 87.5%; Pred. No. 9.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGAGACTC 24  
||||| ||||| ||||| ||||| |||||  
Db 54 AAAAAAACTTGAAGGAGGACGC 77

RESULT 35  
CR477147/c

LOCUS  
DEFINITION  
CR477147 Rat pBluescript Lion Rattus norvegicus cDNA clone  
LI0NP463H10365 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CR477147 315 bp mRNA linear EST 07-JUL-2004  
Rattus norvegicus (Norway rat)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 315)  
Henrich,J., Hermanns,J., Krauz,H., Loebbert,R., Schlueter,T.,  
Schuette,D., Weindel,M., Heil,O., Ebert,L., Neubert,P., Peters,M.,  
Radelof,U., Schneider,D. and Korn,B.  
Rat ArrayTAG cDNA  
Unpublished (2004)

TITLE  
JOURNAL  
COMMENT

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Email: [www.rzpd.de](http://www.rzpd.de)  
RZPD; LI0NP463H10365.  
RZPDLIB;

Rat ArrayTAG cDNA  
<http://www.rzpd.de/cgi-bin/products/showlib.pl.cgi?response=libNo=463> Contact: Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 100  
Fax: +49 30 32639 111  
[www.rzpd.de](http://www.rzpd.de)

This clone is available royalty-free from RZPD;  
contact RZPD ([clone@rzpd.de](mailto:clone@rzpd.de)) for further information. Seq primer:

RP: CAGGAACACAGCTATGAC.  
Location/Qualifiers  
1..315  
/organism="Rattus norvegicus"  
/mol\_type="mrna"  
/db\_xref="taxon:10116"  
/clone="LI0NP463H10365"  
/lab\_host="DH10B"  
/clone\_lib="Rat pBluescript Lion"

FEATURES  
source

ORIGIN

Query Match 76.8%; Score 19.2; DB 7; Length 315;  
Best Local Similarity 87.5%; Pred. No. 9.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGAAGGAGACTCA 25  
||||| ||||| ||||| ||||| |||||  
Db 217 AAAGACACTTGCAGGGTGACTCA 194

RESULT 36  
CO150776

LOCUS  
DEFINITION  
CO150776 Aspergillus flavus Normalized cDNA Expression Library  
EST825829 Aspergillus flavus cDNA clone NAGEK12 5', end similar to (Q7S1X9)  
Aspergillus flavus protein, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CO150776 323 bp mRNA linear EST 17-JUN-2004  
Aspergillus flavus  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; Mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 323)  
Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.  
Aspergillus flavus expressed sequence tags for identification of  
genes with putative roles in aflatoxin contamination of crops  
PEMS Microbiol. Lett. (2004) In press  
Contact: Yu J  
Food and Feed Safety Research Unit  
USDA/ARS, Southern Regional Research Center  
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA  
Tel: 504 286 4405  
Fax: 504 286 4419  
Email: [jiuyu@rrrc.ars.usda.gov](mailto:jiuyu@rrrc.ars.usda.gov)  
Contact Dr. Yu at USDA/ARS SRRC ([jiuyu@rrrc.ars.usda.gov](mailto:jiuyu@rrrc.ars.usda.gov)) for clone  
information  
PCR Primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source

Location/Qualifiers  
1..323  
/organism="Aspergillus flavus"  
/mol\_type="mrna"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAGEK12"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="E. coli DH10B T1 resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression  
Library"  
/note="Vector: pBlueScript (SK+) (Stratagene), antibiotic  
selection marker: Carbenicillin; Site 1: NotI, at the 5  
prime end; Site 2: EcoRI, at the 3 prime end; This  
normalized cDNA expression library was constructed using a  
mixture of mycelial cells grown under eight different  
medium conditions and harvested at 5 time points (18, 24,  
48, 72, 96 hours). The poly-A sequence was trimmed off  
before ligating to vector."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 7; Length 323;  
 Best Local Similarity 87.5%; Pred. No. 9.3e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGAGACTCA 25  
 |||||  
 Db 115 AAAAACTTGGGAAGGAGACTCA 138

## RESULT 37

BI284068/c

LOCUS BI284068 354 bp mRNA linear EST 19-JUL-2001  
 DEFINITION UI-R-DA0-bzd-h-04-0-UI.81 UI-R-DA0 Rattus norvegicus cDNA clone

UI-R-DA0-bzd-h-04-0-UI 3', mRNA sequence.

## ACCESSION

VERSION BI284068

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

1 (bases 1 to 354)

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@iowa.edu

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 non-normalized rat eye library cDNA Library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com)

Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

source

1..354

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-DA0-bzd-h-04-0-UI"

/dev\_stage="ADULT"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-DA0"

/note="Vector: p7713D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DA0  
 library is a non-normalized library constructed from rat  
 salivary gland tissue. For a detailed description of the  
 library from which this clone was derived, please visit  
 our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
 been previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)

TAG TISSUE=rat eye

TAG\_LIB=UI-R-DA0

TAG\_SEQ=CAGCC

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 354;  
 Best Local Similarity 87.5%; Pred. No. 9.4e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGAGACTCA 25  
 |||||  
 Db 232 AAGAGACTTGGCAAGGGTGACTCA 209

## RESULT 38

BI633808/c

LOCUS BI633808 368 bp mRNA linear EST 12-AUG-2003

DEFINITION BX633808 pBluescript Lion Mus musculus cDNA clone LIONp462D0812 3',

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 368)

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Ina Rolf

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; LIONp462D0812.

RZPDLIB;

Mouse ArrayTAG cDNA (LION)

<http://www.rzpd.de/cgi-bin/products/showLib.pl.cgi?response?libNo=4>

62 Contact: Ina Rolf

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

[www.rzpd.de](http://www.rzpd.de)

This clone is available royalty-free from RZPD;

contact RZPD ([clone@rzpd.de](mailto:clone@rzpd.de)) for further information. Seq primer:

RP: CAGGAACAGCTATGAC.

FEATURES

source

1..368

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="LIONp462D0812"

/lab\_host="DH10B"

/clone\_lib="pBluescript Lion"

ORIGIN

Query Match 76.8%; Score 19.2; DB 5; Length 368;

Best Local Similarity 87.5%; Pred. No. 9.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTTGGGAAGGAGACTCA 25  
 |||||  
 Db 217 AAGAGACTTGGCAAGGGTGACTCA 194

## RESULT 39

AK006516

LOCUS AK006516 375 bp mRNA linear HTC 03-APR-2004

DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:1700029022 product:nerve growth factor, gamma, full

insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

HTC; CAP trapper.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;





Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

**TITLE**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL**  
Nature 420, 563-573 (2002)

**PUBLISHED**  
12466851

**COMMENT**  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Konno,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.

**Direct Submission**  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

**FEATURES**  
source  
1. 375  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1700029022"  
/sex="male"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone\_lib="RIKEN full-length enriched, adult male testis"  
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGGAGAGCGCGCCGATTAATTCGAGTTAATTAATATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."

**ORIGIN**  
Query Match 76.8%; Score 19.2; DB 5; Length 375;  
Best Local Similarity 87.5%; Pred. No. 9.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 2 AAAAACAACCTTGAAGGAGACTCA 25  
|||||  
**Db** 159 AAAGACACTTGAAGGGTGACTCA 182  
|||||

**RESULT 41**  
BI139766/c  
**LOCUS**  
IP1.45.A02.b1.A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,  
mRNA sequence.  
**DEFINITION**  
BI139766  
**ACCESSION**  
VERSION BI139766.1 GI:14592209  
**KEYWORDS**  
SOURCE EST.  
**ORGANISM**  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
**REFERENCE**  
1 (bases 1 to 388)  
**AUTHORS**  
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and  
Pratt,L.H.  
**TITLE**  
An EST database from Sorghum: developing preanthesis pannicles  
**JOURNAL**  
Unpublished (2001)  
**COMMENT**  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for high quality sequence is  
20. Three-prime sequences, which are obtained with PolyTMix or T7  
sequencing primer, are presented as the reverse complement.  
Seq primer: JEN REV  
High quality sequence stop: 385  
POLYA=No.

**FEATURES**  
Location/Qualifiers  
1. 388  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultiivar="BTx623"  
/db\_xref="taxon:4558"  
/clone\_lib="Immature pannicle 1 (IP1)"  
/note="Organ: Developing preanthesis pannicles; Vector:  
pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI;  
Site 2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."

**ORIGIN**  
Query Match 76.8%; Score 19.2; DB 2; Length 388;  
Best Local Similarity 87.5%; Pred. No. 9.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 2 AAAAACAACCTTGAAGGAGACTCA 25  
|||||  
**Db** 363 AAAAACAACCTTGAAGGAGACTAA 340  
|||||

**RESULT 42**  
BI284146/c  
**LOCUS**  
UI-R-DA0-bzg-g-04-0-UI.s1 UI-R-DA0 Rattus norvegicus cDNA clone  
**DEFINITION**  
UI-R-DA0-bzg-g-04-0-UI 3', mRNA sequence.  
**ACCESSION**  
VERSION BI284146.1 GI:14936554  
**KEYWORDS**  
SOURCE EST.  
**ORGANISM**  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;



REFERENCE AUTHORS TITLE	Sciurognathi; Muroidea; Muridae; Murinae; Rattus. 1 (bases 1 to 391) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL PUBMED COMMENT	Genome Res. 6 (9), 791-806 (1996) 8889548 Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized rat eye library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) Seq primer: M13 Forward POLYA=Yes.
FEATURES source	Location/Qualifiers 1..391 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague-Dawley" /db_xref="taxon:10116" /clone="UI-R-DA0-bzg-g-04-0-UI" /dev_stage="ADULT" /lab_host="DH10B (Life Technologies)" /clone_lib="UI-R-DA0" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DA0 library is a non-normalized library constructed from rat salivary gland tissue. For a detailed description of the library from which this clone was derived, please visit our web site at <a href="http://rategen.uiowa.edu">rategen.uiowa.edu</a> . The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996) TAG_TISSUE=rat eye TAG_LIB=UI-R-DA0 TAG_SEQ=CAGCC"
ORIGIN	Query Match 76.8%; Score 19.2; DB 2; Length 391; Best Local Similarity 87.5%; Pred. No. 9.6e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  QY 2 AAAAACTTGGAGGGAGACTCA 25       Db 232 AAAGACACTTGCAGGGTGACTCA 209 
RESULT 43 BI139768/c LOCUS	IP1_45_A05_b1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA, mRNA sequence. EST 03-JUL-2001
DEFINITION	BI139768 441 bp mRNA linear EST 03-JUL-2001
ACCESSION	BI139768
VERSION	BI139768.1 GI:14592211
KEYWORDS	EST.
SOURCE	Sorghum bicolor (sorghum)
ORGANISM	Sorghum bicolor Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Focaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE AUTHORS	1 (bases 1 to 441) Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt,L.H.
TITLE	An EST database from Sorghum: developing preanthesis pannicles
JOURNAL COMMENT	Unpublished (2001) Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTwix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 431 POLYA=No.
FEATURES source	Location/Qualifiers 1..441 /organism="Sorghum bicolor" /mol_type="mRNA" /cultivar="BTx623" /db_xref="taxon:4558" /clone_lib="Immature pannicle 1 (IP1)" /note="Organ: Developing preanthesis pannicles; Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
ORIGIN	Query Match 76.8%; Score 19.2; DB 2; Length 441; Best Local Similarity 87.5%; Pred. No. 9.8e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  QY 2 AAAAACTTGGAGGGAGACTCA 25       Db 363 AAAAACTTGCAGAGGAGACTAA 340 
RESULT 44 AQ228531/c LOCUS	AQ228531 473 bp DNA linear GSS 26-SEP-1998
DEFINITION	HS 2023 B2 G01_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2023 Col=2 Row=N, genomic survey sequence.
ACCESSION	AQ228531
VERSION	AQ228531.1 GI:3653760
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
REFERENCE AUTHORS	1 (bases 1 to 473) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL PUBMED COMMENT	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999) 10449764 Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 2023 row: N column: 2 Class: BAC ends High quality sequence stop: 473. Location/Qualifiers 1..473

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2023 Col=2 Row=N"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"

ORIGIN
Query Match          76.8%; Score 19.2; DB 9; Length 473;
Best Local Similarity 87.5%; Pred. No. 9.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACAACCTTGGGAAGGAGACTCA 25
|||||
Db 90 AAGAACACTTGAAGGAGAGTCA 67
|||||

RESULT 45
CE564697
LOCUS
DEFINITION
tigr-gss-dog-1700032751217 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION
CE564697
VERSION
CE564697.1 GI:36881478
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE
1 (bases 1 to 495)
Kirckness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Deicher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1998-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..495
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match          76.8%; Score 19.2; DB 10; Length 495;
Best Local Similarity 87.5%; Pred. No. 9.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACAACCTTGGGAAGGAGACTCA 25
|||||
Db 243 AAAAATACTTGAAGAGAGGCTCA 266
|||||

RESULT 46
AA925291/c
LOCUS
DEFINITION
AA925291
UI-R-AI-ee-h-08-0-UI.s1 UI-R-AI Rattus norvegicus cDNA clone
UI-R-AI-ee-h-08-0-UI 3, similar to gi|205029|gb|M11563|RATKALPS Rat
submaxillary gland PS kallikrein mRNA, complete cds, mRNA sequence.

```

```

ACCESSION
AA925291
VERSION
AA925291.1 GI:4236482
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 498)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
On Apr 21, 1998 this sequence version replaced gi:3072427.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult spleen library. cDNA library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1771334
Seq primer: M13 Forward
POLYA=NO.
FEATURES
Location/Qualifiers
source
1..498
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AI-ee-h-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-AI"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AI
library is a subtracted library derived from the UI-R-A0
library. The UI-R-A0 library consisted of a mixture of
individually tagged normalized libraries constructed from
rat placenta, adult lung, brain, liver, kidney, heart,
spleen, ovary, and muscle. The tag is a string of 3-5'
nucleotides present between the Not I site and the
oligo-dr track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-AI) was constructed as follows: PCR
amplified cDNA inserts from a pool of approximately 3,840
UI-R-A0 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-A0
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-AI library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)"

```

## ORIGIN

```

Query Match          76.8%; Score 19.2; DB 1; Length 498;
Best Local Similarity 87.5%; Pred. No. 9.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACAACCTTGGGAAGGAGACTCA 25
|||||
Db 231 AAAGACACTTGAAGGAGTCA 208
|||||

```



## Coordinated Laboratory for Computational Genomics

University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

## FEATURES

Location/Qualifiers  
1. .605

/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DA0-byj-e-03-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-DA0"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DA0 library is a non-normalized library constructed from rat salivary gland tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_TISSUE=rat eye  
TAG\_LIB=UI-R-DA0  
TAG\_SEQ=CAGCC

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 605;  
Best Local Similarity 87.5%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAACACTTGGAGGGAGACTCA 25

Db 229 AAAGACACTTGCAGGGTGACTCA 206

## RESULT 50

BI279445/c

LOCUS

DEFINITION BI279445 608 bp mRNA linear EST 19-JUL-2001  
UI-R-DA0-by1-e-05-0-UI.s1 UI-R-DA0 Rattus norvegicus cDNA clone

UI-R-DA0-by1-e-05-0-UI 3', mRNA sequence.

ACCESSION BI279445

VERSION BI279445.1

KEYWORDS GI:14927259

SOURCE EST.

ORGANISM Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Rattus.

1 (bases 1 to 608)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

CONTACT: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

## FEATURES

Location/Qualifiers  
1. .608

/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DA0-by1-e-05-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-DA0"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-DA0 library is a non-normalized library constructed from rat salivary gland tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_TISSUE=rat eye  
TAG\_LIB=UI-R-DA0  
TAG\_SEQ=CAGCC

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 608;  
Best Local Similarity 87.5%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAACACTTGGAGGGAGACTCA 25

Db 232 AAAGACACTTGCAGGGTGACTCA 209

Search completed: February 3, 2006, 22:02:04

Job time : 2960.67 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttgaggagactca 25

-Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6 COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6 COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.2	72.8	601	3	US-09-949-016-28002
C 2	18.2	72.8	601	3	US-09-949-016-54169
C 3	18.2	72.8	601	3	US-09-949-016-142990
C 4	18.2	72.8	2092	3	US-09-023-655-936
C 5	18.2	72.8	5203	3	US-09-949-016-1562
C 6	18.2	72.8	6721	3	US-09-949-016-518
C 7	18.2	72.8	47858	3	US-09-949-016-14965
C 8	18.2	72.8	120213	3	US-09-949-016-13304
C 9	18.2	72.8	120217	3	US-09-949-016-12260
C 10	18.2	72.8	265038	3	US-09-949-016-15779
C 11	18.2	72.8	421118	3	US-09-949-016-16297
C 12	17.8	71.2	601	3	US-09-949-016-46494
C 13	17.8	71.2	601	3	US-09-949-016-46495
C 14	17.8	71.2	601	3	US-09-949-016-113510
C 15	17.8	71.2	601	3	US-09-949-016-113511
C 16	17.8	71.2	65745	3	US-09-949-016-12591
C 17	17.8	71.2	165998	3	US-09-949-016-15871
C 18	17.8	71.2	186734	3	US-09-949-016-24
C 19	17.8	71.2	193689	3	US-09-949-016-14870
C 20	17.8	71.2	193689	3	US-09-949-016-12350
C 21	17.8	71.2	193689	3	US-09-949-016-13088
C 22	17.8	71.2	197496	3	US-09-877-177A-10
C 23	17.6	70.4	120	3	US-09-513-999C-17420
C 24	17.6	70.4	299	3	US-08-906-769-92
C 25	17.6	70.4	299	3	US-08-906-769-92
C 26	17.6	70.4	299	3	US-08-817-795-92
C 27	17.6	70.4	299	3	US-08-639-075A-92
C 28	17.6	70.4	299	3	US-09-012-431-92
C 29	17.6	70.4	299	3	US-09-012-692-92
C 30	17.6	70.4	299	3	US-08-906-613-92
C 31	17.6	70.4	299	6	PCT-US95-14442A-92
C 32	17.6	70.4	436	3	US-08-906-769-122
C 33	17.6	70.4	436	3	US-08-906-616-122
C 34	17.6	70.4	436	3	US-08-639-075A-122
C 35	17.6	70.4	436	3	US-09-012-431-122
C 36	17.6	70.4	436	3	US-09-032-215-19
C 37	17.6	70.4	436	3	US-09-012-692-122
C 38	17.6	70.4	436	3	US-08-906-613-122
C 39	17.6	70.4	601	3	US-09-949-016-168761
C 40	17.6	70.4	643	3	US-09-513-999C-10285
C 41	17.6	70.4	674	3	US-08-881-094-13
C 42	17.6	70.4	1152	3	US-09-032-215-24
C 43	17.6	70.4	1152	3	US-09-032-215-25
C 44	17.6	70.4	1303	3	US-09-032-215-21
C 45	17.6	70.4	1303	3	US-09-032-215-23
C 46	17.6	70.4	2627	3	US-09-023-655-188
C 47	17.6	70.4	4029	3	US-09-620-312D-201
C 48	17.6	70.4	13615	3	US-09-949-016-15643
C 49	17.6	70.4	26434	3	US-09-949-016-12338
C 50	17.6	70.4	26434	3	US-09-949-016-17324
C 51	17.6	70.4	45842	3	US-09-949-016-12550
C 52	17.6	70.4	45842	3	US-09-949-016-17327
C 53	17.6	70.4	49003	3	US-09-949-016-16265
C 54	17.6	70.4	157032	3	US-09-949-016-16502
C 55	17.6	70.4	1230025	3	US-09-198-452A-1
C 56	17.6	70.4	1230230	3	US-09-438-185A-1
C 57	17.4	69.6	601	3	US-09-949-016-115272
C 58	17.4	69.6	601	3	US-09-949-016-14542
C 59	17.4	69.6	44166	3	US-09-949-016-15829
C 60	17.4	69.6	54180	3	US-09-949-016-14894
C 61	17.2	68.8	153642	3	US-09-949-016-12174
C 62	17.2	68.8	153643	3	US-09-949-016-15635
C 63	17.2	68.8	314798	3	US-09-949-016-13539
C 64	17.2	68.8	601	3	US-09-949-016-22297
C 65	17.2	68.8	601	3	US-09-949-016-22298
C 66	17.2	68.8	601	3	US-09-949-016-25754
C 67	17.2	68.8	601	3	US-09-949-016-57461
C 68	17.2	68.8	601	3	US-09-949-016-57462
C 69	17.2	68.8	601	3	US-09-949-016-64781
C 70	17.2	68.8	601	3	US-09-949-016-64782
C 71	17.2	68.8	601	3	US-09-949-016-108602
C 72	17.2	68.8	601	3	US-09-949-016-157680
C 73	17.2	68.8	601	3	US-09-949-016-172240
C 74	17.2	68.8	1095	2	US-08-700-607-4
C 75	17.2	68.8	1766	3	US-09-149-476-254
C 76	17.2	68.8	2262	3	US-09-949-016-2988
C 77	17.2	68.8	2664	3	US-09-149-476-255
C 78	17.2	68.8	13906	3	US-09-949-016-14730
C 79	17.2	68.8	13906	3	US-09-949-016-14730
C 80	17.2	68.8	22973	3	US-09-949-016-13644
C 81	17.2	68.8	44988	3	US-09-949-016-16354
C 82	17.2	68.8	50595	3	US-09-949-016-11955
C 83	17.2	68.8	50595	3	US-09-949-016-13410
C 84	17.2	68.8	73788	3	US-09-949-016-12358
C 85	17.2	68.8	87594	3	US-09-949-016-12135
C 86	17.2	68.8	87611	3	US-09-949-016-16139
C 87	17.2	68.8	102008	3	US-09-949-016-16617
C 88	17.2	68.8	105050	3	US-09-949-016-15953
C 89	17.2	68.8	236341	3	US-09-949-016-13978
C 90	17.2	68.8	300402	3	US-09-949-016-13632
C 91	17.2	68.8	317366	3	US-09-949-016-16001
C 92	16.8	67.2	439	3	US-09-513-999C-10832
C 93	16.8	67.2	1524	3	US-10-104-047-1828
C 94	16.8	67.2	2356	3	US-10-104-047-632
C 95	16.8	67.2	2536	3	US-10-104-047-943
C 96	16.8	67.2	3133	3	US-10-104-047-943
C 97	16.8	67.2	3152	3	US-09-710-279-4056

98 16.6 66.4 601 3 US-09-949-016-26792 Sequence 26792, A  
99 16.6 66.4 601 3 US-09-949-016-32735 Sequence 32735, A  
100 16.6 66.4 601 3 US-09-949-016-32736 Sequence 32736, A  
101 16.6 66.4 601 3 US-09-949-016-42795 Sequence 42795, A  
102 16.6 66.4 601 3 US-09-949-016-42796 Sequence 42796, A  
103 16.6 66.4 601 3 US-09-949-016-50585 Sequence 50585, A  
104 16.6 66.4 601 3 US-09-949-016-69121 Sequence 69121, A  
105 16.6 66.4 601 3 US-09-949-016-142348 Sequence 142348, A  
106 16.6 66.4 601 3 US-09-949-016-142349 Sequence 142349, A  
107 16.6 66.4 1558 3 US-09-949-016-764 Sequence 764, App  
108 16.6 66.4 1558 3 US-09-949-016-4025 Sequence 4025, App  
109 16.6 66.4 1709 3 US-09-561-077C-10 Sequence 10, Appl  
110 16.6 66.4 1709 3 US-09-221-014-10 Sequence 10, Appl  
111 16.6 66.4 1776 3 US-09-561-077C-17 Sequence 17, Appl  
112 16.6 66.4 1776 3 US-09-221-014-17 Sequence 17, Appl  
113 16.6 66.4 2319 3 US-09-561-077C-15 Sequence 15, Appl  
114 16.6 66.4 2324 3 US-09-221-014-15 Sequence 15, Appl  
115 16.6 66.4 2324 3 US-09-724-864-21 Sequence 21, Appl  
116 16.6 66.4 3551 3 US-09-561-077C-16 Sequence 16, Appl  
117 16.6 66.4 3551 3 US-09-221-014-16 Sequence 16, Appl  
118 16.6 66.4 7113 3 US-09-561-077C-26 Sequence 26, Appl  
119 16.6 66.4 7113 3 US-09-221-014-26 Sequence 26, Appl  
120 16.6 66.4 19728 3 US-09-949-016-12506 Sequence 12506, A  
121 16.6 66.4 19728 3 US-09-949-016-15767 Sequence 15767, A  
122 16.6 66.4 41380 3 US-09-949-016-15128 Sequence 15128, A  
123 16.6 66.4 42235 3 US-09-949-016-12558 Sequence 12558, A  
124 16.6 66.4 56832 3 US-09-949-016-12576 Sequence 12576, A  
125 16.6 66.4 58844 3 US-09-949-016-13769 Sequence 13769, A  
126 16.6 66.4 59648 3 US-09-949-016-13139 Sequence 13139, A  
127 16.6 66.4 109038 3 US-09-949-016-12199 Sequence 12199, A  
128 16.6 66.4 139257 3 US-09-920-671-11 Sequence 11, Appl  
129 16.6 66.4 162465 3 US-09-949-016-14264 Sequence 14264, A  
130 16.6 66.4 192700 3 US-09-949-016-11820 Sequence 11820, A  
131 16.6 66.4 192704 3 US-09-949-016-17182 Sequence 17182, A  
132 16.6 66.4 198942 3 US-09-949-016-13209 Sequence 13209, A  
133 16.6 66.4 276237 3 US-09-949-016-17504 Sequence 17504, A  
134 16.6 65.6 263 3 US-09-205-258-656 Sequence 656, App  
135 16.4 65.6 263 3 US-09-205-258-657 Sequence 657, App  
136 16.4 65.6 263 3 US-09-205-258-658 Sequence 658, App  
137 16.4 65.6 263 3 US-10-004-860-656 Sequence 656, App  
138 16.4 65.6 263 3 US-10-004-860-657 Sequence 657, App  
139 16.4 65.6 263 3 US-10-004-860-658 Sequence 658, App  
140 16.4 65.6 568 3 US-09-247-155-136 Sequence 136, App  
141 16.4 65.6 568 3 US-09-303-190-136 Sequence 136, App  
142 16.4 65.6 588 3 US-09-205-258-64 Sequence 64, Appl  
143 16.4 65.6 588 3 US-10-004-860-64 Sequence 64, Appl  
144 16.4 65.6 1304 3 US-09-799-451-57 Sequence 57, Appl  
145 16.2 64.8 176 3 US-09-513-999C-12329 Sequence 12329, A  
146 16.2 64.8 362 3 US-09-640-211A-1637 Sequence 1637, App  
147 16.2 64.8 389 3 US-09-640-211A-1587 Sequence 1587, App  
148 16.2 64.8 465 3 US-09-621-976-14430 Sequence 14430, A  
149 16.2 64.8 601 3 US-09-949-016-23811 Sequence 23811, A  
150 16.2 64.8 601 3 US-09-949-016-32333 Sequence 32333, A

ALIGNMENTS

RESULT 1  
US-09-949-016-28002/c  
; Sequence 28002, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

RESULT 2  
US-09-949-016-54169/c  
; Sequence 54169, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

RESULT 3  
US-09-949-016-142990  
; Sequence 142990, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03



```
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-142990

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACAATTGGAGGAGACT 24
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 530 AAAAACAATTGGAGGAGACT 552

RESULT 4
US-09-023-655-936
; Sequence 936, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; TITLE OF INVENTION: Jeffrey J. Sellhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 853-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 936:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2092 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1381145
US-09-023-655-936

Query Match          72.8%; Score 18.2; DB 3; Length 2092;
Best Local Similarity 87.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACAATTGGAGGAGACT 23
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 724 AAAAACAATTGGAGGAGACT 746

RESULT 5
US-09-949-016-1562
; Sequence 1562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1562
; LENGTH: 5203
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1562

Query Match          72.8%; Score 18.2; DB 3; Length 5203;
Best Local Similarity 87.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACAATTGGAGGAGACT 23
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1613 AAAAACAATTGGAGGAGACT 1635

RESULT 6
US-09-949-016-518
; Sequence 518, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 518
; LENGTH: 6721
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-518

Query Match          72.8%; Score 18.2; DB 3; Length 6721;
Best Local Similarity 87.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACAATTGGAGGAGACT 23
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1618 AAAAACAATTGGAGGAGACT 1640

RESULT 7
US-09-949-016-14965/c
; Sequence 14965, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```



```
RESULT 11
US-09-949-016-16297/c
; Sequence 16297, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16297
; LENGTH: 421118
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(421118)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16297

Query Match 72.8%; Score 18.2; DB 3; Length 421118;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGGAAGGAGACT 23
Db 59077 AACTACACTTGGGAAGGAGAAAT 59055

RESULT 12
US-09-949-016-46494/c
; Sequence 46494, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46494
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-46494

Query Match 71.2%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGGAAGGAGGA 21
Db 539 ATAAACACTTAGGAAGGAGGA 519

RESULT 13
US-09-949-016-113511/c
; Sequence 113511, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
US-09-949-016-46495/c
; Sequence 46495, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46495
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-46495

Query Match 71.2%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGGAAGGAGGA 21
Db 379 ATAAACACTTAGGAAGGAGGA 359

RESULT 14
US-09-949-016-113510/c
; Sequence 113510, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113510
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-113510

Query Match 71.2%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGGAAGGAGGA 21
Db 539 ATAAACACTTAGGAAGGAGGA 519

RESULT 15
US-09-949-016-113511/c
; Sequence 113511, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113511
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(65744)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-113511

Query Match 71.2%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGA 21
| | | | | | | | | | | | | | | | | | | | |
Db 379 ATAAACACTAGAGGGAGA 359

RESULT 16
US-09-949-016-12591
; Sequence 12591, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12591
; LENGTH: 65744
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(65744)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12591

Query Match 71.2%; Score 17.8; DB 3; Length 65744;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGA 21
| | | | | | | | | | | | | | | | | | | | |
Db 34052 AAAAAACAGTTGGAAGGAAGA 34072

RESULT 17
US-09-949-016-15871
; Sequence 15871, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15871
; LENGTH: 65745
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(65745)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15871

Query Match 71.2%; Score 17.8; DB 3; Length 65745;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGAGA 21
| | | | | | | | | | | | | | | | | | | | |
Db 34052 AAAAAACAGTTGGAAGGAAGA 34072

RESULT 18
US-09-676-610B-24
; Sequence 24, Application US/09676610B
; Patent No. 644465
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
; FILE REFERENCE: RTS-0138
; CURRENT APPLICATION NUMBER: US/09/676,610B
; CURRENT FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 182
; SEQ ID NO 24
; LENGTH: 169998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1208)...(1472)
; NAME/KEY: intron
; LOCATION: (1473)...(124390)
; NAME/KEY: exon
; LOCATION: (124391)...(124544)
; NAME/KEY: intron
; LOCATION: (124545)...(125409)
; NAME/KEY: exon
; LOCATION: (125410)...(125595)
; NAME/KEY: intron
; LOCATION: (125596)...(128711)
; NAME/KEY: exon
; LOCATION: (128712)...(128848)
; NAME/KEY: intron
; LOCATION: (128849)...(133400)
; NAME/KEY: exon
; LOCATION: (133401)...(133469)
; NAME/KEY: intron
; LOCATION: (133470)...(134652)
; NAME/KEY: exon
; LOCATION: (134653)...(134773)
; NAME/KEY: intron
; LOCATION: (134774)...(136116)
```

/ NAME/KEY: exon  
/ LOCATION: (136117)...(136261)  
/ NAME/KEY: intron  
/ LOCATION: (136262)...(137936)  
/ NAME/KEY: exon  
/ LOCATION: (137937)...(138053)  
/ NAME/KEY: intron  
/ LOCATION: (138054)...(138637)  
/ NAME/KEY: exon  
/ LOCATION: (138638)...(138766)  
/ NAME/KEY: intron  
/ LOCATION: (138767)...(138864)  
/ NAME/KEY: exon  
/ LOCATION: (138865)...(138940)  
/ NAME/KEY: intron  
/ LOCATION: (138941)...(139765)  
/ NAME/KEY: exon  
/ LOCATION: (139766)...(139860)  
/ NAME/KEY: intron  
/ LOCATION: (139861)...(142245)  
/ NAME/KEY: exon  
/ LOCATION: (142246)...(142445)  
/ NAME/KEY: intron  
/ LOCATION: (142446)...(143605)  
/ NAME/KEY: exon  
/ LOCATION: (143606)...(143738)  
/ NAME/KEY: intron  
/ LOCATION: (143739)...(145938)  
/ NAME/KEY: exon  
/ LOCATION: (145839)...(145931)  
/ NAME/KEY: intron  
/ LOCATION: (145932)...(147385)  
/ NAME/KEY: exon  
/ LOCATION: (147386)...(147544)  
/ NAME/KEY: intron  
/ LOCATION: (147545)...(153274)  
/ NAME/KEY: exon  
/ LOCATION: (153275)...(153321)  
/ NAME/KEY: intron  
/ LOCATION: (153322)...(155088)  
/ NAME/KEY: exon  
/ LOCATION: (155089)...(155231)  
/ NAME/KEY: intron  
/ LOCATION: (155232)...(156025)  
/ NAME/KEY: exon  
/ LOCATION: (156026)...(156151)  
/ NAME/KEY: intron  
/ LOCATION: (156152)...(156826)  
/ NAME/KEY: exon  
/ LOCATION: (156827)...(156928)  
/ NAME/KEY: intron  
/ LOCATION: (156929)...(163399)  
/ NAME/KEY: exon  
/ LOCATION: (163400)...(163586)  
US-09-676-610B-24

Query Match 71.2%; Score 17.8; DB 3; Length 169998;  
Best Local Similarity 90.5%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAACAACCTTGGAGGGGAGAC 22  
Db 104105 AAAAACAACCTTGGAGGGGAGAC 104125

RESULT 19  
US-09-949-016-14870/c  
/ Sequence 14870, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTER, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

/ FILE REFERENCE: CL001307  
/ CURRENT APPLICATION NUMBER: US/09/949,016  
/ CURRENT FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20  
/ PRIOR APPLICATION NUMBER: 60/237,768  
/ PRIOR FILING DATE: 2000-10-03  
/ PRIOR APPLICATION NUMBER: 60/231,498  
/ PRIOR FILING DATE: 2000-09-08  
/ NUMBER OF SEQ ID NOS: 207012  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 14870  
/ LENGTH: 186734  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-949-016-14870

Query Match 71.2%; Score 17.8; DB 3; Length 186734;  
Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACAACCTTGGAGGGGAGA 21  
Db 23486 ATAAACAACCTAGGAGGGGAGA 23466

RESULT 20  
US-09-949-016-12350/c  
/ Sequence 12350, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTER, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
/ FILE REFERENCE: CL001307  
/ CURRENT APPLICATION NUMBER: US/09/949,016  
/ CURRENT FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20  
/ PRIOR APPLICATION NUMBER: 60/237,768  
/ PRIOR FILING DATE: 2000-10-03  
/ PRIOR APPLICATION NUMBER: 60/231,498  
/ PRIOR FILING DATE: 2000-09-08  
/ NUMBER OF SEQ ID NOS: 207012  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 12350  
/ LENGTH: 193689  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-949-016-12350

Query Match 71.2%; Score 17.8; DB 3; Length 193689;  
Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACAACCTTGGAGGGGAGA 21  
Db 23486 ATAAACAACCTAGGAGGGGAGA 23466

RESULT 21  
US-09-949-016-13088/c  
/ Sequence 13088, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTER, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
/ FILE REFERENCE: CL001307  
/ CURRENT APPLICATION NUMBER: US/09/949,016  
/ CURRENT FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13088  
; LENGTH: 193689  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13088

Query Match 71.2%; Score 17.8; DB 3; Length 193689;  
Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGGGAGA 21  
Db 23486 ATAAACACTAGGAAGGGGAGA 23466

RESULT 22  
US-09-877-177A-10  
; Sequence 10, Application US/09877177A  
; Patent No. 6582919  
; GENERAL INFORMATION:  
; APPLICANT: K. Danenberg  
; TITLE OF INVENTION: Method of determining Epidermal Growth  
; TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression  
; TITLE OF INVENTION: and Correlation of Levels Thereof With Survival  
; FILE REFERENCE: 11220/120  
; CURRENT APPLICATION NUMBER: US/09/877,177A  
; CURRENT FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 197496  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-877-177A-10

Query Match 71.2%; Score 17.8; DB 3; Length 197496;  
Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAACACTTGGAGGGGAGAC 22  
Db 112105 AAAAGACTTGGAGGGGAGAC 112125

RESULT 23  
US-09-513-999C-17420/c  
; Sequence 17420, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 17420  
; LENGTH: 120  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-513-999C-17420

Query Match 70.4%; Score 17.6; DB 3; Length 120;  
Best Local Similarity 83.3%; Pred. No. 67;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGGGAGACTC 24  
Db 114 AAAAAAACTTGTAAAGTTAGACTC 91

RESULT 24  
US-08-906-769-92  
; Sequence 92, Application US/08906769  
; Patent No. 6077687  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Silver, Gary  
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES AND USES THEREOF  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/906,769  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/639,075  
; FILING DATE: 24-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..299  
; OTHER INFORMATION: /note= "At pos. bp 178/179, change  
; OTHER INFORMATION: G/C to V/Y. At pos. aa 59, substitue Xaa."  
US-08-906-769-92

Query Match 70.4%; Score 17.6; DB 3; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAACACTTGGAGGGGAGACTCA 25  
Db 267 AAAGACACTTGCAGGGGAGATTCA 290

RESULT 25  
US-08-906-616-92  
; Sequence 92, Application US/08906616  
; Patent No. 6121035  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Silver, Gary  
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/906,616  
; FILING DATE: 05-AUG-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..299  
; OTHER INFORMATION: /note= "At pos. bp 178/179, change  
; G/C to V/Y. At pos. aa 59, substitute Xaa."  
US-08-906-616-92

Query Match 70.4%; Score 17.6; DB 3; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGGAGACTCA 25  
||| ||||| ||||| ||||| |||||  
Db 267 AAAGACACTTGGCCAGGGAGATTCA 290

RESULT 26  
US-08-817-795-92  
; Sequence 92, Application US/08817795  
; Patent No. 6139840  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Heath, Andrew W.  
; APPLICANT: Yamaka, Miles Yamanaka  
; APPLICANT: Arfsten, Ann  
; APPLICANT: Dale, Beverly

; APPLICANT: Stiegler, Gary  
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND  
; PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA  
; INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; MOLECULES, AND USES THEREOF  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,795  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/14442  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gary J. Connell  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..299  
; NAME/KEY: Xaa = any amino acid  
; LOCATION: 59  
US-08-817-795-92

Query Match 70.4%; Score 17.6; DB 3; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGGAGACTCA 25  
||| ||||| ||||| ||||| |||||  
Db 267 AAAGACACTTGGCCAGGGAGATTCA 290

RESULT 27  
US-08-639-075A-92  
; Sequence 92, Application US/08639075A  
; Patent No. 6150125  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Silver, Gary  
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; MOLECULES AND USES THEREOF  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh



STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 92:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..299  
OTHER INFORMATION: /note= "At pos. bp 178/179, change  
G/C to V/Y. At pos. aa 59, substitute Xaa."  
US-08-639-075A.92

Query Match : 70.4%; Score 17.6; DB 3; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTTGAAGGAGACTCA 25  
Db 267 AAAGACACTTGCAGGAGATTCA 290

## RESULT 28

US-09-012-431-92  
Sequence 92, Application US/09012431  
Patent No. 6180383

## GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/012,431

FILING DATE: 23-Jan-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 92:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..299  
OTHER INFORMATION: /note= "At pos. bp 178/179, change  
G/C to V/Y. At pos. aa 59, substitute Xaa."  
US-09-012-431-92

Query Match : 70.4%; Score 17.6; DB 3; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTTGAAGGAGACTCA 25  
Db 267 AAAGACACTTGCAGGAGATTCA 290

## RESULT 29

US-09-012-692-92

Sequence 92, Application US/09012692

Patent No. 6214579

## GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/012,692

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA: US 08/639,075

APPLICATION NUMBER: US 08/639,075

FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

```

; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..299
; OTHER INFORMATION: /note= "At pos. bp 178/179, change
; OTHER INFORMATION: G/C to V/Y. At pos. aa 59, substitute Xaa."
US-09-012-692-92

Query Match 70.4%; Score 17.6; DB 3; Length 299;
Best Local Similarity 83.3%; Pred. No. 80;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGAGACTCA 25
Db 267 AAAGACACTTGCAGGAGATTCA 290

RESULT 31
PCT-US95-14442A-92
; Sequence 92, Application PC/TUS9514442A
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Arfsten, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..299
; NAME/KEY: Xaa = any amino acid
; LOCATION: 59
PCT-US95-14442A-92
```

```

; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..299
; OTHER INFORMATION: /note= "At pos. bp 178/179, change
; OTHER INFORMATION: G/C to V/Y. At pos. aa 59, substitute Xaa."
US-09-012-692-92

Query Match 70.4%; Score 17.6; DB 3; Length 299;
Best Local Similarity 83.3%; Pred. No. 80;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGAGACTCA 25
Db 267 AAAGACACTTGCAGGAGATTCA 290

RESULT 30
US-08-906-613-92
; Sequence 92, Application US/08906613
; Patent No. 6232096
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

Query Match 70.4%; Score 17.6; DB 6; Length 299;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGAAGGAGACTCA 25  
|||||  
Db 267 AAAGACACTTGCAGGAGATTCA 290  
|||||

RESULT 32  
US-08-906-769-122  
; Sequence 122, Application US/08906769  
; Patent No. 6077687  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Silver, Gary  
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/906,769  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/639,075  
; FILING DATE: 24-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 122:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 436 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2..436  
; OTHER INFORMATION: /note= "at pos. bp 301, change A to  
; OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to  
; OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,  
; OTHER INFORMATION: substitute xaa."  
US-08-906-769-122

Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGAAGGAGACTCA 25  
|||||  
Db 254 AAAGACACTTGCAGGAGATTCA 277  
|||||

RESULT 34  
US-08-639-075A-122  
; Sequence 122, Application US/08639075A  
; Patent No. 6150125  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley

Db 254 AAAGACACTTGCAGGAGATTCA 277

RESULT 33  
US-08-906-616-122  
; Sequence 122, Application US/08906616  
; Patent No. 6121035  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Stiegler, Gary  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Silver, Gary  
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/906,616  
; FILING DATE: 05-AUG-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Connell, Gary J.  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 122:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 436 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2..436  
; OTHER INFORMATION: /note= "at pos. bp 301, change A to  
; OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to  
; OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,  
; OTHER INFORMATION: substitute xaa."  
US-08-906-616-122

Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGAAGGAGACTCA 25  
|||||  
Db 254 AAAGACACTTGCAGGAGATTCA 277  
|||||

RESULT 34  
US-08-639-075A-122  
; Sequence 122, Application US/08639075A  
; Patent No. 6150125  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..436  
OTHER INFORMATION: /note= "At pos. bp 301, change A to  
W; at pos. bp 342, change C to Y; at pos. bp 397, change C to  
OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,  
OTHER INFORMATION: substitute Xaa."  
US-08-639-075A-122  
Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 AAAAACACTTGGAGGGAGACTCA 25  
Db 254 AAAGACACTTGCAGGGAGATTCA 277  
RESULT 35  
US-09-012-431-122  
Sequence 122, Application US/09012431  
Patent No. 6180383  
GENERAL INFORMATION:  
APPLICANT: Griewe, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,431  
FILING DATE: 23-Jan-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..436  
OTHER INFORMATION: /note= "At pos. bp 301, change A to  
W; at pos. bp 342, change C to Y; at pos. bp 397, change C to Y;  
at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,  
substitute Xaa."  
US-09-012-431-122  
Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 AAAAACACTTGGAGGGAGACTCA 25  
Db 254 AAAGACACTTGCAGGGAGATTCA 277  
RESULT 36  
US-09-032-215-19  
Sequence 19, Application US/09032215  
Patent No. 6204010  
GENERAL INFORMATION:  
APPLICANT: Stiegler, Gary L.  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,215

/ FILING DATE: 27-FEB-1998  
/ CLASSIFICATION: 536  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Connell, Gary J.  
/ REGISTRATION NUMBER: 32,020  
/ REFERENCE/DOCKET NUMBER: 2618-25-C6  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (303) 863-9700  
/ TELEFAX: (303) 863-0223  
/ INFORMATION FOR SEQ ID NO: 13:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 436 nucleotides  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: cDNA  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: 1..436  
/ FEATURE:  
/ NAME/KEY: W = A  
/ LOCATION: 301  
/ FEATURE:  
/ NAME/KEY: Y = C  
/ LOCATION: 342, 397  
/ FEATURE:  
/ NAME/KEY: S = C  
/ LOCATION: 431  
/ FEATURE:  
/ NAME/KEY: Xaa = Unknown  
/ LOCATION: 100, 114, 144  
/ US-09-032-215-19

Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACACTTGAAGGAGACTCA 25  
||| ||||| ||||| |||||  
Db 254 AAAGACACTTGCAGGAGATTCA 277

RESULT 37  
US-09-012-692-122  
/ Sequence 122, Application US/09012692  
/ Patent No. 6214579  
/ GENERAL INFORMATION:  
/ APPLICANT: Grieve, Robert B.  
/ APPLICANT: Rushlow, Keith E.  
/ APPLICANT: Wu Hunter, Shirley  
/ APPLICANT: Frank, Glenn R.  
/ APPLICANT: Stiegler, Gary  
/ APPLICANT: Gaines, Patrick J.  
/ APPLICANT: Silver, Gary  
/ TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
/ NUMBER OF SEQUENCES: 190  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Sheridan Ross & McIntosh  
/ STREET: 1700 Lincoln Street, Suite 3500  
/ CITY: Denver  
/ STATE: Colorado  
/ COUNTRY: USA  
/ ZIP: 80203  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09012692  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/639,075  
/ FILING DATE: 24-APR-1996  
/ ATTORNEY/AGENT INFORMATION:

/ APPLICATION NUMBER: US/09/012.692  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/639,075  
/ FILING DATE: 24-APR-1996  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Connell, Gary J.  
/ REGISTRATION NUMBER: 32,020  
/ REFERENCE/DOCKET NUMBER: 2618-25-C2  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (303) 863-9700  
/ TELEFAX: (303) 863-0223  
/ INFORMATION FOR SEQ ID NO: 122:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 436 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: cDNA  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: 2..436  
/ OTHER INFORMATION: /note= "At pos. bp 301, change A to  
/ OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to  
/ OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,  
/ OTHER INFORMATION: substitute Xaa."  
/ US-09-012-692-122

Query Match 70.4%; Score 17.6; DB 3; Length 436;  
Best Local Similarity 83.3%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAACACTTGAAGGAGACTCA 25  
||| ||||| ||||| |||||  
Db 254 AAAGACACTTGCAGGAGATTCA 277

RESULT 38  
US-08-906-613-122  
/ Sequence 122, Application US/08906613  
/ Patent No. 6232096  
/ GENERAL INFORMATION:  
/ APPLICANT: Grieve, Robert B.  
/ APPLICANT: Rushlow, Keith E.  
/ APPLICANT: Wu Hunter, Shirley  
/ APPLICANT: Frank, Glenn R.  
/ APPLICANT: Stiegler, Gary  
/ APPLICANT: Gaines, Patrick J.  
/ APPLICANT: Silver, Gary  
/ TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
/ NUMBER OF SEQUENCES: 190  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Sheridan Ross & McIntosh  
/ STREET: 1700 Lincoln Street, Suite 3500  
/ CITY: Denver  
/ STATE: Colorado  
/ COUNTRY: USA  
/ ZIP: 80203  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/906,613  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/639,075  
/ FILING DATE: 24-APR-1996  
/ ATTORNEY/AGENT INFORMATION:

```
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 122:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..436
; OTHER INFORMATION: /note= "at pos. bp 301, change A to
; OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
; OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
; OTHER INFORMATION: substitute Xaa."
;
US-08-906-613-122

Query Match 70.4%; Score 17.6; DB 3; Length 436;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTTGGAGGGAGACTCA 25
||| ||||| ||||| |||
Db 254 AAAGACACTTGCAGGGAGATTCA 277

RESULT 39
US-09-949-016-168761/c
; Sequence 168761, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168761
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-168761

Query Match 70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 92;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTTGGAGGGAGACTC 24
||| ||||| ||||| |||
Db 47 AAAAAATAATTGTAATGGAGACTC 24

RESULT 40
US-09-513-999C-10285/c
; Sequence 10285, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duciert, A.
; APPLICANT: Giordano, J.Y.
```

```
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59,US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 10285
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 274
; OTHER INFORMATION: w=a or t
;
US-09-513-999C-10285

Query Match 70.4%; Score 17.6; DB 3; Length 643;
Best Local Similarity 83.3%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTTGGAGGGAGACTC 24
||| ||||| ||||| |||
Db 632 AAAAACTTGGAGGGGTACCC 609

RESULT 41
US-08-881-094-13
; Sequence 13, Application US/08881094A
; Patent No. 6022739
; GENERAL INFORMATION:
; APPLICANT: Ryan, Clarence A
; APPLICANT: Pearce, Gregory L
; APPLICANT: McGurl, Barry F
; TITLE OF INVENTION: Systemin
; FILE REFERENCE: 7555-000001CPB
; CURRENT APPLICATION NUMBER: US/08/881,094A
; CURRENT FILING DATE: 1997-07-09
; EARLIER APPLICATION NUMBER: 08/308,887
; EARLIER FILING DATE: 1994-09-19
; EARLIER APPLICATION NUMBER: PCT/US93/02428
; EARLIER FILING DATE: 1993-03-18
; EARLIER APPLICATION NUMBER: 07/885,412
; EARLIER FILING DATE: 1992-03-19
; EARLIER APPLICATION NUMBER: 07/528,956
; EARLIER FILING DATE: 1990-05-25
; EARLIER APPLICATION NUMBER: PCT/US91/03685
; EARLIER FILING DATE: 1991-05-24
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 674
; TYPE: DNA
; ORGANISM: Solanum tuberosum
;
US-08-881-094-13

Query Match 70.4%; Score 17.6; DB 3; Length 674;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTTGGAGGGAGACTC 24
||| ||||| ||||| |||
Db 139 AAAAAATAATTGAAAGGAGACTC 162

RESULT 42
US-09-032-215-24
; Sequence 24, Application US/09032215
; Patent No. 6204010
; GENERAL INFORMATION:
; APPLICANT: Stiegler, Gary L.
```

```

RESULT 44
US-09-032-215-21
; Sequence 21, Application US/09032215
; Patent No. 6204010
; GENERAL INFORMATION:
; APPLICANT: Stiegler, Gary L.
; APPLICANT: Gaines, Patrick J.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,215  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1303 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 149..1300  
FEATURE:  
NAME/KEY: W = A or T/U  
LOCATION: 1165  
FEATURE:  
NAME/KEY: Y = C or T/U  
LOCATION: 1206,1261  
FEATURE:  
NAME/KEY: S = C or G  
LOCATION: 1295  
FEATURE:  
NAME/KEY: Xaa = Unknown  
LOCATION: 339, 353, 383  
US-09-032-215-21

Query Match 70.4%; Score 17.6; DB 3; Length 1303;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGACTTGGAGGAGACTCA 25  
DB 1118 AAAGACACTTGCAGGAGATTCA 1141

RESULT 45  
US-09-032-215-23/c  
Sequence 23, Application US/09032215  
Patent No. 6204010  
GENERAL INFORMATION:  
APPLICANT: Stiegler, Gary L.  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,215

FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1303 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: R = A or G  
LOCATION: 43, 98  
FEATURE:  
NAME/KEY: W = A or T/U  
LOCATION: 139  
FEATURE:  
NAME/KEY: S = C or G  
LOCATION: 9  
US-09-032-215-23

Query Match 70.4%; Score 17.6; DB 3; Length 1303;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGACTTGGAGGAGACTCA 25  
DB 186 AAAGACACTTGCAGGAGATTCA 163

RESULT 46  
US-09-023-655-188  
Sequence 188, Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HERewith  
CLASSIFICATION:  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071

```
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGNOT01
; CLONE: 060309
US-09-023-655-188

Query Match 70.4%; Score 17.6; DB 3; Length 2627;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACTC 24
DB 405 AAACAACACTTGGAAATTC 428

RESULT 47
US-09-620-312D-201
; Sequence 201, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 201
; LENGTH: 4029
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (316)..(1842)
US-09-620-312D-201

Query Match 70.4%; Score 17.6; DB 3; Length 4029;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACTC 24
DB 1803 AAACAACACTTGGAAATTC 1826

; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGNOT01
; CLONE: 060309
US-09-023-655-188

Query Match 70.4%; Score 17.6; DB 3; Length 2627;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACTC 24
DB 405 AAACAACACTTGGAAATTC 428

RESULT 48
US-09-949-016-15643/c
; Sequence 15643, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15643
; LENGTH: 13615
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-15643

Query Match 70.4%; Score 17.6; DB 3; Length 13615;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACTC 24
DB 4021 AAAAACACTTGAAGAGTCACTC 3998

RESULT 49
US-09-949-016-12338/c
; Sequence 12338, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12338
; LENGTH: 26434
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-12338

Query Match 70.4%; Score 17.6; DB 3; Length 26434;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACTCA 25
DB 24563 AAACGACTTGAAGAGATCTTA 24540

RESULT 50
US-09-949-016-17324/c
; Sequence 17324, Application US/09949016
```

; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 17324  
; LENGTH: 26434  
; TYPE: DNA  
; ORGANISM: Human  
; US-09-949-016-17324

Query Match 70.4%; Score 17.6; DB 3; Length 26434;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAGCACTTGGAGGAGACTCA 25  
Db 24563 AAAAAGCACTTGGAGGAGACTCA 24540

Search completed: February 3, 2006, 16:32:12  
Job time : 87.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttgaggagactca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	US-10-719-900-10
2	23.4	93.6	25	8	US-10-719-900-9
3	20.2	80.8	600	9	US-10-972-079-48833
C 4	19.8	79.2	1043	4	US-09-925-065A-91527
C 5	19.2	76.8	441	7	US-10-767-701-6654
C 6	19.2	76.8	873	7	US-10-152-319A-1905
C 7	18.8	75.2	1078	7	US-10-424-599-69851
C 8	18.6	74.4	833	4	US-09-925-065A-264666
C 9	18.6	74.4	897	4	US-09-925-065A-708512
10	18.6	74.4	1268	7	US-10-043-160-22
C 11	18.6	74.4	2511	3	US-09-823-245A-367
C 12	18.6	74.4	9210	3	US-09-764-877-3622
C 13	18.6	74.4	9210	3	US-09-764-877-3635
C 14	18.6	74.4	9210	6	US-10-242-515-3622
C 15	18.6	74.4	9210	6	US-10-242-515-3635
C 16	18.6	74.4	62169	7	US-10-322-696-157
C 17	18.4	73.6	114280	8	US-10-719-993-6808
C 18	18.2	72.8	188	6	US-10-029-386-23105
C 19	18.2	72.8	434	5	US-10-027-632-183294
C 20	18.2	72.8	434	6	US-10-027-632-183294
C 21	18.2	72.8	582	4	US-09-925-065A-548984
C 22	18.2	72.8	582	4	US-09-925-065A-548985
C 23	18.2	72.8	590	6	US-10-029-386-9405

24	18.2	72.8	786	3	US-09-938-842A-3478	Sequence 3478, Ap
25	18.2	72.8	786	3	US-09-938-842A-3478	Sequence 3478, Ap
C 26	18.2	72.8	865	5	US-10-198-846-7215	Sequence 7215, Ap
27	18.2	72.8	1260	6	US-10-369-493-33506	Sequence 33506, A
28	18.2	72.8	2092	3	US-09-968-007A-422	Sequence 422, App
29	18.2	72.8	2092	7	US-10-641-643-936	Sequence 936, App
30	18.2	72.8	2092	9	US-10-843-644A-6892	Sequence 6892, Ap
C 31	18.2	72.8	2306	6	US-10-369-493-46352	Sequence 46352, A
32	18.2	72.8	6721	7	US-10-693-999-19	Sequence 19, Appl
33	18.2	72.8	6787	9	US-10-887-553A-586	Sequence 586, App
34	17.8	71.2	368	8	US-10-425-115-116569	Sequence 116569, App
35	17.8	71.2	490	4	US-09-925-065A-423916	Sequence 423916, App
36	17.8	71.2	490	4	US-09-925-065A-423917	Sequence 423917, App
37	17.8	71.2	525	9	US-10-487-901-4349	Sequence 4349, Ap
C 38	17.8	71.2	619	4	US-09-925-065A-313637	Sequence 313637, App
C 39	17.8	71.2	619	4	US-09-925-065A-313638	Sequence 313638, App
C 40	17.8	71.2	619	4	US-09-925-065A-313639	Sequence 313639, App
41	17.8	71.2	631	4	US-09-925-065A-724596	Sequence 724596, App
42	17.8	71.2	631	4	US-09-925-065A-724597	Sequence 724597, App
43	17.8	71.2	631	4	US-09-925-065A-724598	Sequence 724598, App
C 44	17.8	71.2	5347	5	US-10-205-823-199	Sequence 199, App
C 45	17.8	71.2	5347	10	US-11-051-454-199	Sequence 199, Appl
46	17.8	71.2	169998	6	US-10-380-931-24	Sequence 24, Appl
47	17.8	71.2	197496	3	US-09-877-177-10	Sequence 10, Appl
48	17.8	71.2	197496	6	US-10-426-836-10	Sequence 10, Appl
49	17.8	71.2	209083	9	US-10-461-862-74	Sequence 74, Appl
C 50	17.8	71.2	334462	9	US-10-496-011-1	Sequence 1, Appl
51	17.6	70.4	25	10	US-11-036-317-158180	Sequence 158180, Appl
52	17.6	70.4	25	10	US-11-036-317-158180	Sequence 158180, Appl
C 53	17.6	70.4	208	7	US-10-424-599-127517	Sequence 127517, Appl
54	17.6	70.4	215	7	US-10-424-599-120024	Sequence 120024, Appl
55	17.6	70.4	287	7	US-10-424-599-120024	Sequence 120024, Appl
C 56	17.6	70.4	300	9	US-10-779-543-6493	Sequence 6493, Ap
57	17.6	70.4	311	7	US-10-242-535A-35405	Sequence 35405, A
58	17.6	70.4	311	7	US-10-085-783A-35405	Sequence 35405, A
C 59	17.6	70.4	389	9	US-10-779-543-11595	Sequence 11595, A
C 60	17.6	70.4	456	8	US-10-674-124A-7165	Sequence 7165, Ap
C 61	17.6	70.4	525	5	US-10-425-115-30106	Sequence 30106, A
C 62	17.6	70.4	555	5	US-10-066-543-2547	Sequence 2547, Ap
C 63	17.6	70.4	561	4	US-09-925-065A-171539	Sequence 171539, Appl
64	17.6	70.4	563	3	US-09-969-034-1995	Sequence 1995, Ap
C 65	17.6	70.4	567	3	US-09-998-598-1866	Sequence 1866, Ap
C 66	17.6	70.4	574	4	US-09-925-065A-171540	Sequence 171540, Appl
C 67	17.6	70.4	585	5	US-10-027-632-187263	Sequence 187263, Appl
C 68	17.6	70.4	585	5	US-10-027-632-187264	Sequence 187264, Appl
C 69	17.6	70.4	585	5	US-10-027-632-187265	Sequence 187265, Appl
C 70	17.6	70.4	585	5	US-10-027-632-187266	Sequence 187266, Appl
C 71	17.6	70.4	585	6	US-10-027-632-187263	Sequence 187263, Appl
C 72	17.6	70.4	585	6	US-10-027-632-187264	Sequence 187264, Appl
C 73	17.6	70.4	585	6	US-10-027-632-187265	Sequence 187265, Appl
C 74	17.6	70.4	585	6	US-10-027-632-187266	Sequence 187266, Appl
75	17.6	70.4	586	4	US-09-925-065A-408057	Sequence 408057, Appl
76	17.6	70.4	786	7	US-09-925-065A-677975	Sequence 677975, Appl
C 77	17.6	70.4	786	7	US-10-152-319A-1512	Sequence 1512, Ap
C 78	17.6	70.4	1110	6	US-10-312-273-276	Sequence 276, App
C 79	17.6	70.4	1407	8	US-10-425-115-30110	Sequence 30110, A
C 80	17.6	70.4	1607	5	US-10-357-930-23368	Sequence 23368, A
C 81	17.6	70.4	1966	8	US-10-116-802-358	Sequence 358, App
C 82	17.6	70.4	2081	6	US-10-062-674-1369	Sequence 1369, App
C 83	17.6	70.4	2477	6	US-10-094-749-744	Sequence 744, App
C 84	17.6	70.4	2516	6	US-10-094-749-900	Sequence 900, App
85	17.6	70.4	2627	7	US-10-641-643-188	Sequence 188, App
C 86	17.6	70.4	3856	5	US-10-084-817-116	Sequence 116, App
C 87	17.6	70.4	3878	9	US-10-956-157-4653	Sequence 4653, Ap
C 88	17.6	70.4	4015	5	US-10-106-698-1018	Sequence 1018, Ap
C 89	17.6	70.4	4029	5	US-10-037-270-201	Sequence 201, App
90	17.6	70.4	4029	9	US-10-117-722-201	Sequence 201, App
91	17.6	70.4	4029	9	US-10-122-851-201	Sequence 201, App
C 92	17.6	70.4	5184	3	US-09-764-891-5785	Sequence 5785, Ap
C 93	17.6	70.4	94752	6	US-10-408-168-34	Sequence 34, Appl
94	17.6	70.4	140040	7	US-10-275-762-69	Sequence 69, Appl
95	17.6	70.4	156318	8	US-10-741-600-17574	Sequence 17574, A
96	17.6	70.4	166472	8	US-10-741-600-17788	Sequence 17788, A

```
97 17.6 70.4 174703 5 US-10-087-192-1336
98 17.6 70.4 247544 7 US-10-322-696-55
99 17.6 70.4 312477 7 US-10-317-88A-12
100 17.6 70.4 1230025 6 US-10-289-762-1
101 17.4 69.6 578 4 US-09-925-065A-330456
102 17.4 69.6 578 4 US-09-925-065A-330457
103 17.4 69.6 597 9 US-10-972-079-24527
104 17.4 69.6 600 9 US-10-972-079-24526
105 17.4 69.6 606 4 US-09-925-065A-330455
106 17.4 69.6 668 4 US-09-925-065A-647756
107 17.4 69.6 2682 3 US-09-938-842A-1842
108 17.4 69.6 2682 3 US-09-938-842A-1842
109 17.4 69.6 109559 7 US-10-322-281-137
110 17.2 68.8 300 9 US-10-779-543-3255
111 17.2 68.8 340 8 US-10-425-113-74382
112 17.2 68.8 402 9 US-10-779-543-11739
113 17.2 68.8 493 8 US-10-357-930-7162
114 17.2 68.8 525 4 US-09-925-065A-326219
115 17.2 68.8 589 4 US-09-925-065A-166948
116 17.2 68.8 596 4 US-09-925-065A-408056
117 17.2 68.8 599 9 US-10-972-079-84232
118 17.2 68.8 607 4 US-09-925-065A-727194
119 17.2 68.8 609 4 US-09-925-065A-646991
120 17.2 68.8 611 4 US-09-925-065A-154947
121 17.2 68.8 618 4 US-09-925-065A-836275
122 17.2 68.8 626 4 US-09-925-065A-554232
123 17.2 68.8 631 5 US-10-027-632-185105
124 17.2 68.8 631 6 US-10-027-632-185105
125 17.2 68.8 642 4 US-09-925-065A-780072
126 17.2 68.8 962 4 US-09-925-065A-705133
127 17.2 68.8 1082 4 US-09-925-065A-763
128 17.2 68.8 1082 4 US-09-925-065A-764
129 17.2 68.8 1082 4 US-09-925-065A-49681
130 17.2 68.8 1727 3 US-09-764-864-1611
131 17.2 68.8 2793 3 US-09-938-842A-394
132 17.2 68.8 2793 7 US-09-938-842A-394
133 17.2 68.8 2793 7 US-10-225-067-1
134 17.2 68.8 2793 7 US-10-374-780A-2561
135 17.2 68.8 3302 4 US-09-925-065A-553893
136 17.2 68.8 3526 7 US-10-437-963-60029
137 17.2 68.8 10368 10 US-11-097-143-5041
138 17.2 68.8 42863 6 US-10-017-161-1017
139 17.2 68.8 42863 6 US-10-292-798-865
140 17.2 68.8 325446 8 US-10-719-993-6824
141 17.2 68.8 374849 5 US-10-087-192-1627
142 17.2 68.8 383432 9 US-10-737-082-34
143 17.2 68.8 383432 9 US-10-765-790-34
144 17.2 68.8 1790242 8 US-10-719-993-6940
145 17 68.0 25 7 US-10-719-956-1799
146 17 68.0 25 7 US-10-719-956-1800
147 17 68.0 104 3 US-09-918-995-18212
148 17 68.0 112 3 US-09-918-995-18215
149 17 68.0 186 7 US-10-424-599-66385
150 17 68.0 201 7 US-10-741-601-870
```

## ALIGNMENTS

## RESULT 1

US-10-719-900-10

```
; Sequence 10, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10719,900
; PRIOR FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; NUMBER OF SEQ ID NOS: 2002 11 20
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

```
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-10

Query Match 100.0%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25
|||||
Db 1 AAAAAACACTTGGAGGAGACTCA 25
|||||

RESULT 2
US-10-719-900-9
; Sequence 9, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 9
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-9

Query Match 93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.76;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25
|||||
Db 1 AAAAAACACTTGGAGGAGACTCA 25
|||||

RESULT 3
US-10-972-079-48833
; Sequence 48833, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BSEI
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 48833
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894294418_1
US-10-972-079-48833

Query Match 80.8%; Score 20.2; DB 9; Length 600;
Best Local Similarity 88.0%; Pred. No. 35;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 AAAAAACACTTGGGAAGGAGACTCA 25  
|||||  
Db 10 AAAAAACACTTGGGAAGGAGACTCA 34

RESULT 4  
US-09-925-065A-91527/c  
; Sequence 91527, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 91527  
; LENGTH: 1043  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-91527

Query Match 79.2%; Score 19.8; DB 4; Length 1043;  
Best Local Similarity 91.3%; Pred. No. 58;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGGAAGGAGACT 23  
|||||  
Db 367 AAAAAACACTTGGGAAGGAGACT 345

RESULT 5  
US-10-767-701-6654/c  
; Sequence 6654, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 6654  
; LENGTH: 441  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(441)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS50873\_1  
US-10-767-701-6654

Query Match 76.8%; Score 19.2; DB 7; Length 441;  
Best Local Similarity 87.5%; Pred. No. 98;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGGGAAGGAGACTCA 25  
|||||  
Db 363 AAAAAACACTTGGGAAGGAGACTAA 340

RESULT 6  
US-10-152-319A-1905  
; Sequence 1905, Application US/10152319A  
; Publication No. US20040072160A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Higgs, Brandon  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5089-US  
; CURRENT APPLICATION NUMBER: US/10/152,319A  
; CURRENT FILING DATE: 2002-05-22  
; PRIOR APPLICATION NUMBER: US 60/292,335  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/297,523  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,925  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,810  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/303,807  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/303,808  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/315,047  
; PRIOR FILING DATE: 2001-08-28  
; PRIOR APPLICATION NUMBER: US 60/324,928  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US 60/330,867  
; PRIOR FILING DATE: 2001-11-01  
; PRIOR APPLICATION NUMBER: US 60/330,462  
; PRIOR FILING DATE: 2001-10-22  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 2221  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1905  
; LENGTH: 873  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. NM\_031523  
US-10-152-319A-1905

Query Match 76.8%; Score 19.2; DB 7; Length 873;  
Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGGGAAGGAGACTCA 25  
|||||  
Db 658 AAAAAACACTTGGGAAGGAGACTCA 681

RESULT 7  
US-10-424-599-69851/c  
; Sequence 69851, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
US-10-424-599-69851/c



```
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 69851
; LENGTH: 1078
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34088C.1
US-10-424-599-69851

Query Match          75.2%; Score 18.8; DB 7; Length 1078;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAACACTTGGAGGAGACTC 24
   |||||
Db 779 AAAACACTTGGAGGAGACTC 758

RESULT 8
US-09-925-065A-264666/c
; Sequence 264666, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264666
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-264666

Query Match          74.4%; Score 18.6; DB 4; Length 597;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGAGACTCA 25
   |||||
Db 306 AGAAGACACTTCGAAGGAGACTCA 282

RESULT 9
US-09-925-065A-708512
; Sequence 708512, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20

; CURRENT APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-20
; CURRENT APPLICATION NUMBER: 09/554,911
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/FR98/02501
; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: FR 97/14669
; PRIOR FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 1268
; TYPE: DNA
; ORGANISM: Gallus sp.
US-10-043-160-22

Query Match          74.4%; Score 18.6; DB 7; Length 1268;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGAGACTCA 25
   |||||
Db 992 AAAAAATAGTGGAGGAGAACTCA 1016

RESULT 11
US-09-823-245A-367/c
; Sequence 367, Application US/09823245A
; Publication No. US20020039760A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fectel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
```

APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6401  
; CURRENT APPLICATION NUMBER: US/09/823,245A  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/194,941  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 367  
; LENGTH: 2511  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-823-245A-367

Query Match 74.4%; Score 18.6; DB 3; Length 2511;  
Best Local Similarity 84.0%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25  
|||||  
Db 1682 AAAAAACACTTGGAGGAGACTCA 1658

RESULT 12  
US-09-764-877-3622/c  
; Sequence 3622, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3622  
; LENGTH: 9210  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-3622

Query Match 74.4%; Score 18.6; DB 3; Length 9210;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25  
|||||  
Db 7719 AAAAAACACTTGGAGGAGACTCA 7695

RESULT 13  
US-09-764-877-3635/c  
; Sequence 3635, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3635  
; LENGTH: 9210  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-3635

Query Match 74.4%; Score 18.6; DB 3; Length 9210;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACACTTGGAGGAGACTCA 25  
|||||  
Db 7719 AAAAAACACTTGGAGGAGACTCA 7695

RESULT 14  
US-10-242-515-3622/c  
; Sequence 3622, Application US/10242515  
; Publication No. US20040009488A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005C1  
; CURRENT APPLICATION NUMBER: US/10/242,515  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 09/764,877  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3622  
; LENGTH: 9210  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-242-515-3622

Query Match 74.4%; Score 18.6; DB 6; Length 9210;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25  
|||||  
Db 7719 AAAAAACACTTGGAGGAGACTCA 7695

RESULT 15  
US-10-242-515-3635/c  
; Sequence 3635, Application US/10242515  
; Publication No. US20040009488A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005C1  
; CURRENT APPLICATION NUMBER: US/10/242,515  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 09/764,877  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28

```
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3635
; LENGTH: 9210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-515-3635

Query Match          74.4%; Score 18.6; DB 6; Length 9210;
Best Local Similarity 84.0%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
DB 7719 AAAAAACACTTGGAGGAGACTCA 7695

RESULT 16
US-10-322-696-157/c
; Sequence 157, Application US/10322696
; Publication No. US20040166490A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Malandro, Marc
; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
; FILE REFERENCE: 529452001200
; CURRENT APPLICATION NUMBER: US/10/322,696
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157
; LENGTH: 62169
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(62169)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-696-157

Query Match          74.4%; Score 18.6; DB 7; Length 62169;
Best Local Similarity 84.0%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
DB 7740 AAAGAAACTTGGAGGAGACTCA 7716

RESULT 17
US-10-719-993-6808
; Sequence 6808, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
```

```
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6808
; LENGTH: 114280
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-6808

Query Match          73.6%; Score 18.4; DB 8; Length 114280;
Best Local Similarity 95.0%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAAACACTTGGAGGAGACT 23
    ||||| ||||| ||||| ||||| |||||
DB 82496 AAACAATTGGAGGAGACT 82515

RESULT 18
US-10-029-386-23105/c
; Sequence 23105, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23105
; LENGTH: 188
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: NT HIT: gi4503164, EVALUE 1.00e-79
; OTHER INFORMATION: SWISSPROT HIT: Q9JLV5, EVALUE 3.00e-26
; OTHER INFORMATION: EST_HUMAN HIT: BF038331.1, EVALUE 2.00e-79
US-10-029-386-23105

Query Match          72.8%; Score 18.2; DB 6; Length 188;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGACT 23
    ||||| ||||| ||||| ||||| |||||
DB 86 AAACAACACTTGGCAAGGAGACT 64

RESULT 19
US-10-027-632-183294/c
; Sequence 183294, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
```

```

, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 183294
, LENGTH: 434
, TYPE: DNA
, ORGANISM: Human
US-10-027-632-183294

```

Query Match 72.8%; Score 18.2; DB 5; Length 434;  
Best Local Similarity 87.0%; Pred. No. 2.8e+02;  
Matches 20: Conservative 0; Mismatches 3; Indels

QY 1 AAAAAACACTTGAAGGGAGACT 23  
||||| ||||| ||||| ||||| ||  
pb 111 AAAAAATACTTGTGAAGGGAGGCT 89

## RESULT 20

US-10-027-632-183294/c  
; Sequence 183294, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Sing

TITLE OF INVENTION: Polymorphisms in the Human Genome

; TITLE OF INVENTION: POLYMER  
; REFERENCE: 109827 138

```

, FILE REFERENCE: 108827.129
, CURRENT APPLICATION NUMBER: US/10/027,632
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 183294
, LENGTH: 434
, TYPE: DNA
, ORGANISM: Human
, US-10-027-632,183294

```

Query Match	72.8%;	Score 18.2;	DB 6;	Length 434;
Best Local Similarity	87.0%;	Pred. No. 2.8e+02;		
Matches	20;	Conservative	0;	Mismatches 3;
				Indels

Qy 1 AAAAAACACTTGAAGCGGAGCT 23  
|||  
Db 111 AAAAAATACCTGTAAAGCGGAGCT 89

## RESULT 21

US-09-925-065A-548984  
; Sequence 548984, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.

```

; TITLE OF INVENTION: Identification and Mapping of Single
;
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;
; FILE REFERENCE: 108927.135
;
; CURRENT APPLICATION NUMBER: US/09/925,065A
;
; CURRENT FILING DATE: 2001-08-08
;
; PRIOR APPLICATION NUMBER: US 60/243,096
;
; PRIOR FILING DATE: 2000-10-24
;
; PRIOR APPLICATION NUMBER: US 60/252,147
;
; PRIOR FILING DATE: 2000-11-20
;
; PRIOR APPLICATION NUMBER: US 60/350,092
;
; PRIOR FILING DATE: 2000-11-30
;
; PRIOR APPLICATION NUMBER: US 60/261,766
;
; PRIOR FILING DATE: 2001-01-16
;
; PRIOR APPLICATION NUMBER: US 60/289,846
;
; PRIOR FILING DATE: 2001-05-09
;
; NUMBER OF SEQ ID NOS: 957086
;
; SOFTWARE: Fast-SEQ for Windows Version 4.0
;
; SEQ ID NO 548984
;
; LENGTH: 582
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; PS-09-925-065A-548984

```

Query Match	72.88:	Score 18.2:	DB 4:	Length 582:
-------------	--------	-------------	-------	-------------

Query Match	Score	Length	Index
Best Local Similarity	87.0%	Pred. No. 2.9e+02;	0.0
Query Match	72.8%	Score 18.4; DP 1;	0.0
Best Local Similarity	87.0%	Pred. No. 2.9e+02;	0.0
Query Match	72.8%	Score 18.4; DP 1;	0.0

**Qy** 1 AAAAAACACTTGGAAAGGAGACT 23  
| | | | | | | | | |  
**Dh** 417 AAAAAACACTTCAGGGGAGACT 439

## RESULT 22

US-09-925-065A-548985  
; Sequence 548985, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.

```

1 APPLICANT: Wang, Xiaohua
2 TITLE OF INVENTION: Identification and Mapping of Single
3 NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME
4 FILE REFERENCE: 108827.135
5 CURRENT APPLICATION NUMBER: US/09/925,065A
6 CURRENT FILING DATE: 2001-08-08
7 PRIOR APPLICATION NUMBER: US 60/243,096
8 PRIOR FILING DATE: 2000-10-24
9 PRIOR APPLICATION NUMBER: US 60/252,147
10 PRIOR FILING DATE: 2000-11-20
11 PRIOR APPLICATION NUMBER: US 60/250,092
12 PRIOR FILING DATE: 2000-11-30
13 PRIOR APPLICATION NUMBER: US 60/261,766
14 PRIOR FILING DATE: 2001-01-16
15 PRIOR APPLICATION NUMBER: US 60/289,846
16 PRIOR FILING DATE: 2001-05-09
17 NUMBER OF SEQ ID NOS: 957086
18 SOFTWARE: FastSEQ for Windows Version 4.0
19 SEQ ID NO 548985
20 LENGTH: 582

```

ORGANISM: Homo sapiens  
UHS-09-925-065A-548985

Query Match	72.8%;	Score 18.2;	DB 4;	Length 582;
Best Local Similarity	87.0%;	Pred. No. 2.9e+02;		
Matches	20.	Mismatches	3;	Indels
Conservative				

QY 1 AAAAAACACTTGGAAAGGAGACT 23  
 |||||  
 417 AAAAAACACTTTCAGGGAGACT 439  
 |||||

## RESULT 23

US-10-029-386-9405/c  
: Sequence 9405. Application US/10029386

US-09-938-842A-3478

Sequence 3478, Application US/09938842A  
Publication No. US20040009476A9  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 3478  
LENGTH: 786  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3478

Query Match 72.8%; Score 18.2; DB 3; Length 786;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACT 24  
|||||

Db 512 ATAAACACTTGGAGAGAGATC 534  
|||||

RESULT 26

US-10-198-846-7215/c

Sequence 7215, Application US/10198846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
TITLE OF INVENTION: THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7215  
LENGTH: 865  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 78, 216, 238, 246, 250, 252, 283, 299, 439, 451, 481, 525,  
LOCATION: 532, 614, 622, 625, 645, 646, 663, 675, 689, 713, 718, 723,  
LOCATION: 729, 739, 740, 756, 766, 771, 794, 810, 816, 820, 829, 836,  
LOCATION: 838, 849, 856, 859, 862  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-7215

Query Match 72.8%; Score 18.2; DB 5; Length 865;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACT 23  
|||||

US-09-938-842A-3478

Sequence 3478, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 3478  
LENGTH: 786  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3478

Query Match 72.8%; Score 18.2; DB 3; Length 786;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACT 24  
|||||

Db 512 ATAAACACTTGGAGAGAGATC 534  
|||||

RESULT 25

US-09-938-842A-3478

Sequence 3478, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 3478  
LENGTH: 786  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3478

Query Match 72.8%; Score 18.2; DB 3; Length 786;  
Best Local Similarity 87.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGAGGGAGACT 24  
|||||

Db 512 ATAAACACTTGGAGAGAGATC 534  
|||||

US-10-029-386-9405

Query Match 72.8%; Score 18.2; DB 6; Length 590;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGGAGACT 23  
|||||

Db 275 AACACACTTGGCAGGAGACT 253  
|||||

Db 575 AAAAAAGACTTGGAGGTAAC 553

RESULT 27

US-10-369-493-33506  
; Sequence 33506, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 33506  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: Desulfotobacterium hafniese  
US-10-369-493-33506

Query Match 72.8%; Score 18.2; DB 6; Length 1260;  
Best Local Similarity 87.0%; Pred. No. 3.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACT 23  
Db 25 AAACAACAGTGGGAAGGTAGACT 47

RESULT 28

US-09-968-007A-422  
; Sequence 422, Application US/09968007A  
; Publication No. US20040115625A1  
; GENERAL INFORMATION:  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signa  
; FILE REFERENCE: 689290-71  
; CURRENT APPLICATION NUMBER: US/09/968,007A  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,172  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,173  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,278  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,294  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,295  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,316  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 1001  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 422  
; LENGTH: 2092  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-968-007A-422

Query Match 72.8%; Score 18.2; DB 3; Length 2092;  
Best Local Similarity 87.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACT 23

Db 724 AAAAAACACTTGGCAAGGAGACT 746

RESULT 29

US-10-641-643-936  
; Sequence 936, Application US/10641643  
; Publication No. US20040077003A1  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Suan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/641,643  
; FILING DATE: 14-Aug-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 936:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2092 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: G1381145  
; SEQUENCE DESCRIPTION: SEQ ID NO: 936 :  
US-10-641-643-936

Query Match 72.8%; Score 18.2; DB 7; Length 2092;  
Best Local Similarity 87.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACT 23  
Db 724 AAAAAACACTTGGCAAGGAGACT 746

RESULT 30

US-10-843-641A-6892  
; Sequence 6892, Application US/10843641A  
; Publication No. US20050064454A1  
; GENERAL INFORMATION:  
; APPLICANT: Avalon Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; TITLE OF INVENTION: Signature Gene Sets  
; FILE REFERENCE: 689290-189  
; CURRENT APPLICATION NUMBER: US/10/843,641A  
; CURRENT FILING DATE: 2004-05-12  
; PRIOR APPLICATION NUMBER: US/09/873,367

```
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,832
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/964,824
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/967,768
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/09/968,007
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 8447
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6892
; LENGTH: 2092
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2092)
; OTHER INFORMATION: n=a,t,g or c
US-10-843-641A-6892

Query Match 72.8%; Score 18.2; DB 9; Length 2092;
Best Local Similarity 87.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGGAGACT 23
DB 724 AAACAACACTTGGCAAGGAGACT 746

RESULT 31
US-10-369-493-46352/c
; Sequence 46352, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46352
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46352

Query Match 72.8%; Score 18.2; DB 6; Length 2306;
Best Local Similarity 87.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACACTTGGAGGGGAGACT 24
DB 1161 ACAACAACACTAGGAGGGGAGGCTC 1139
```

```
RESULT 32
US-10-693-999-19
; Sequence 19, Application US/10693999
; Publication No. US20040137597A1
; GENERAL INFORMATION:
; APPLICANT: Meso Scale Technologies, LLC
; APPLICANT: Davydov, Ilia
; APPLICANT: Kenten, John H.
; APPLICANT: Biebuyck, Hans
; APPLICANT: Oberoi, Pankaj
; TITLE OF INVENTION: THE UBIQUITYLATION OF THESE SUBSTRATES
; FILE REFERENCE: 2528-10 / P14050US0
; CURRENT APPLICATION NUMBER: US/10/693,999
; CURRENT FILING DATE: 2003-10-28
; PRIOR APPLICATION NUMBER: US 60/422,448
; PRIOR FILING DATE: 2002-10-30
; PRIOR APPLICATION NUMBER: US 60/486,529
; PRIOR FILING DATE: 2003-07-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 6721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-693-999-19

Query Match 72.8%; Score 18.2; DB 7; Length 6721;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGGAGACT 23
DB 1618 AAACAACACTTGGCAAGGAGACT 1640

RESULT 33
US-10-887-553A-586
; Sequence 586, Application US/10887553A
; Publication No. US20050085436A1
; GENERAL INFORMATION:
; APPLICANT: Garza, Dan
; APPLICANT: Li, Hao
; TITLE OF INVENTION: Method to treat conditions associated
; FILE REFERENCE: 4-33262
; CURRENT APPLICATION NUMBER: US/10/887,553A
; CURRENT FILING DATE: 2004-07-08
; PRIOR APPLICATION NUMBER: 60/485,883
; PRIOR FILING DATE: 2003-08-07
; NUMBER OF SEQ ID NOS: 1208
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 586
; LENGTH: 6787
; TYPE: DNA
; ORGANISM: human
US-10-887-553A-586

Query Match 72.8%; Score 18.2; DB 9; Length 6787;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGGAGACT 23
DB 1685 AAACAACACTTGGCAAGGAGACT 1707

RESULT 34
US-10-425-115-116569
; Sequence 116569, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
```



```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 116569
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_3779C.1
; US-10-425-115-116569

Query Match      71.2%; Score 17.8; DB 8; Length 368;
Best Local Similarity 90.5%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGA 21
   ||||| ||||| ||||| |||||
Db 168 AAAAAACACTTAGGAAGGGGA 188

RESULT 35
US-09-925-065A-423916
; Sequence 423916, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 423916
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-423916

Query Match      71.2%; Score 17.8; DB 4; Length 490;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGA 21
   ||||| ||||| ||||| |||||
Db 355 AAAAAACACTTGGAGGGAGA 375

RESULT 36
US-09-925-065A-423917
; Sequence 423917, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 423916
; LENGTH: 490
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-423916

Query Match      71.2%; Score 17.8; DB 4; Length 490;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGGAGA 21
   ||||| ||||| ||||| |||||
Db 355 AAAAAACACTTGGAGGGAGA 375

RESULT 37
US-10-487-901-4349
; Sequence 4349, Application US/10487901
; Publication No. US20050091708A1
; GENERAL INFORMATION:
; APPLICANT: Oreido, Jeremiah Vincent
; APPLICANT: McCreery, David
; APPLICANT: Pell, Randy
; APPLICANT: Miller, Barbara
; APPLICANT: Weglarz, Thaddeus
; APPLICANT: Gachotte, Daniel
; APPLICANT: Blakeslee, Beth
; APPLICANT: Larrinua, Ignacio
; APPLICANT: Reddy, Avutu
; APPLICANT: Shukla, Vipula
; APPLICANT: Crosley, Rodney
; TITLE OF INVENTION: Nucleic Acid Compositions Conferring Altered Metabolic Character
; FILE REFERENCE: DOW-08552
; CURRENT APPLICATION NUMBER: US/10/487,901
; CURRENT FILING DATE: 2004-02-26
; NUMBER OF SEQ ID NOS: 7560
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4349
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-487-901-4349

Query Match      71.2%; Score 17.8; DB 9; Length 525;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACACTTGGAGGGAGACTCA 25
   ||||| ||||| ||||| |||||
Db 284 AACACTTGGAGAGACTCA 304

RESULT 38
US-09-925-065A-313637/c
; Sequence 313637, Application US/09925065A
; Publication No. US20050228172A9
```



RESULT 42  
US-09-925-065A-724597  
; Sequence 724597, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 724597  
; LENGTH: 631  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-724597

Query Match 71.2%; Score 17.8; DB 4; Length 631;  
Best Local Similarity 90.5%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGCA 21  
|||||  
Db 284 AAAAAACACTTGGAGGAGCA 304

RESULT 43  
US-09-925-065A-724598  
; Sequence 724598, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 724598  
; LENGTH: 631  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-724598

Query Match 71.2%; Score 17.8; DB 4; Length 631;  
Best Local Similarity 90.5%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGCA 21  
|||||  
Db 284 AAAAAACACTTGGAGGAGCA 304

RESULT 44  
US-10-205-823-199/c  
; Sequence 199, Application US/10205823  
; Publication No. US20030108963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Monsey, Angela M.  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Anderson, Dustin  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER  
; FILE REFERENCE: MRI-044  
; CURRENT APPLICATION NUMBER: US/10/205,823  
; CURRENT FILING DATE: 2002-07-25  
; PRIOR APPLICATION NUMBER: 60/307,982  
; PRIOR FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/314,356  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/325,020  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: 60/341,746  
; PRIOR FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: 60/362,158  
; PRIOR FILING DATE: 2002-03-05  
; NUMBER OF SEQ ID NOS: 455  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 199  
; LENGTH: 5347  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-205-823-199

Query Match 71.2%; Score 17.8; DB 5; Length 5347;  
Best Local Similarity 90.5%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAGGAGCA 21  
|||||  
Db 1038 AAAAAACACTTGGAGGAGCA 1018

RESULT 45  
US-11-051-454-199/c  
; Sequence 199, Application US/11051454  
; Publication No. US20050191673A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Monsey, Angela M.  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Anderson, Dustin  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER

```
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/11/051,454
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199
; LENGTH: 5347
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-051-454-199
```

```
Query Match 71.2%; Score 17.8; DB 10; Length 5347;
Best Local Similarity 90.5%; Pred. No. 6.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAACACTTGGAGGGGAGA 21
||| ||| ||| ||| ||| ||| |||
Db 1038 AAGAAACACATGGAAGGGGAGA 1018
```

```
RESULT 46
US-10-380-931-24
; Sequence 24, Application US/10380931
; Publication No. US20030215944A1
; GENERAL INFORMATION:
; APPLICANT: Isis Pharmaceuticals, Inc.
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
; FILE REFERENCE: RTSP-0187
; CURRENT APPLICATION NUMBER: US/10/380,931
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: 09/676,610
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 182
; SEQ ID NO 24
; LENGTH: 169998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1208)...(1472)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1473)...(124390)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (124391)...(124544)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (124545)...(125409)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (125410)...(125595)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (125596)...(128711)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (128712)...(128848)
```

```
; FEATURE:
; NAME/KEY: intron
; LOCATION: (128849)...(133400)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (133401)...(133469)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (133470)...(134652)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (134653)...(134773)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (134774)...(136116)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (136117)...(136261)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (136262)...(137936)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (137937)...(138053)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (138054)...(138637)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (138638)...(138766)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (138767)...(138864)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (138865)...(138940)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (138941)...(139765)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (139766)...(139860)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (139861)...(142245)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (142246)...(142445)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (142446)...(143605)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (143606)...(143738)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (143739)...(145838)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (145839)...(145931)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (145932)...(147385)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (147386)...(147544)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (147545)...(153274)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (153275)...(153321)
; FEATURE:
```

```
; NAME/KEY: intron
; LOCATION: (153322)...(155088)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (155089)...(155231)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (155232)...(156025)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (156026)...(156151)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (156152)...(156826)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (156827)...(156928)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (156929)...(163399)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (163400)...(163586)
; US-10-380-931-24

Query Match          71.2%; Score 17.8; DB 6; Length 169998;
Best Local Similarity 90.5%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGGAAGGGAGAC 22
||||| ||||| ||||| |||||
Db 104105 AAAAGACTTGGAGGGGAGAC 104125

RESULT 47
US-09-877-177-10
; Sequence 10, Application US/09877177
; Publication No. US20020192652A1
; GENERAL INFORMATION:
; APPLICANT: Peter V. Danenberg et al.
; TITLE OF INVENTION: Method of determining Epidermal Growth
; FILE REFERENCE: 11220/120
; CURRENT APPLICATION NUMBER: US/09/877,177
; CURRENT FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-877-177-10

Query Match          71.2%; Score 17.8; DB 3; Length 197496;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGGAAGGGAGAC 22
||||| ||||| ||||| |||||
Db 112105 AAAAGACTTGGAGGGGAGAC 112125

RESULT 48
US-10-426-836-10
; Sequence 10, Application US/10426836
; Publication No. US20030211530A1
; GENERAL INFORMATION:
; APPLICANT: K. Danenberg
; TITLE OF INVENTION: Method of determining Epidermal Growth
; FILE REFERENCE: 11220/169
; CURRENT APPLICATION NUMBER: US/10/426,836
```

```
; CURRENT FILING DATE: 2003-05-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-426-836-10

Query Match          71.2%; Score 17.8; DB 6; Length 197496;
Best Local Similarity 90.5%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGGAAGGGAGAC 22
||||| ||||| ||||| |||||
Db 112105 AAAAGACTTGGAGGGGAGAC 112125

RESULT 49
US-10-461-862-74
; Sequence 74, Application US/10461862
; Publication No. US20050090434A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 209083
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-461-862-74

Query Match          71.2%; Score 17.8; DB 9; Length 209083;
Best Local Similarity 90.5%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACACTTGGGAAGGGAGAC 22
||||| ||||| ||||| |||||
Db 112983 AAAAGACTTGGAGGGGAGAC 113003

RESULT 50
US-10-496-011-1/c
; Sequence 1, Application US/10496011
; Publication No. US20050118588A1
; GENERAL INFORMATION:
; APPLICANT: BOURGERON, THOMAS
; APPLICANT: JAWAIN, STEPHANE
; APPLICANT: QUACH, HELENE
; APPLICANT: BETANCUR, CATALINA
; APPLICANT: LEBOVER, MARION
; APPLICANT: GILLBERG, CHRISTOPHER
; TITLE OF INVENTION: POLYNUCLEOTIDE AND PROTEIN INVOLVED IN SYNAPTOGENESIS, VARIANTS
; FILE REFERENCE: 253820USXPCT
; CURRENT APPLICATION NUMBER: US/10/496,011
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: PCT/FR02/04134
; PRIOR FILING DATE: 2002-11-28
; PRIOR APPLICATION NUMBER: CA2364106
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 334462
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: exon
; LOCATION: (10670)..(10828)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (83744)..(84513)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (201676)..(201828)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (240359)..(240469)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (312940)..(313125)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (318313)..(319102)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (327018)..(330406)
US-10-496-011-1
```

```
Query Match 71.2%; Score 17.8; DB 9; Length 334462;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAACACTTGGAGGGGAGA 21
||| ||| ||| ||| ||| ||| |||
Db 198779 AAATAACATTGGAGGGGAGA 198759
```

```
Search completed: February 3, 2006, 15:43:41
Job time : 372.556 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-10

Perfect score: 25

Sequence: 1 aaaaacacttgaaggagactca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	76.8	600	7	US-10-750-185-2224
C 2	19.2	76.8	600	7	US-10-750-623-2224
C 3	19.2	76.8	777	8	US-11-136-527-3865
C 4	19.2	76.8	777	8	US-11-136-527-7961
C 5	19.2	76.8	3221	7	US-10-750-185-54511
C 6	19.2	76.8	3221	7	US-10-750-623-54511
C 7	18.8	75.2	3481	7	US-10-750-185-27280
C 8	18.8	75.2	3481	7	US-10-750-623-27280
C 9	17.8	71.2	191331	8	US-11-112-908-20
C 10	17.6	70.4	25	8	US-11-136-527-189320
C 11	17.6	70.4	518	8	US-11-136-527-3680
C 12	17.6	70.4	518	8	US-11-136-527-7776
C 13	17.6	70.4	759	7	US-10-750-185-50114
C 14	17.6	70.4	759	7	US-10-750-623-50114
C 15	17.6	70.4	859	8	US-11-136-527-3681
C 16	17.6	70.4	859	8	US-11-136-527-7777
C 17	17.6	70.4	1582	7	US-10-750-185-42140
C 18	17.6	70.4	1582	7	US-10-750-623-42140
C 19	17.6	70.4	1698	7	US-10-750-185-36667
C 20	17.6	70.4	1698	7	US-10-750-623-36667
C 21	17.2	68.8	1805	7	US-10-750-185-53076
C 22	17.2	68.8	1805	7	US-10-750-623-53076

23	17.2	68.8	2655	7	US-10-750-185-35807	Sequence 35807, A
24	17.2	68.8	2655	7	US-10-750-623-35807	Sequence 35807, A
25	17.2	68.8	186442	8	US-11-121-086-104	Sequence 104, App
26	17	68.0	201	7	US-10-995-561-1528	Sequence 1528, Ap
27	17	68.0	201	7	US-10-995-561-1530	Sequence 1530, Ap
28	17	68.0	201	7	US-10-995-561-1538	Sequence 1538, Ap
29	17	68.0	1047	7	US-10-750-185-42937	Sequence 42937, A
30	17	68.0	1047	7	US-10-750-623-42937	Sequence 42937, A
C 31	17	68.0	1613	7	US-10-750-185-62645	Sequence 62645, A
C 32	17	68.0	1613	7	US-10-750-623-62645	Sequence 62645, A
C 33	17	68.0	1698	7	US-10-750-185-56656	Sequence 56656, A
C 34	17	68.0	1698	7	US-10-750-623-56656	Sequence 56656, A
35	17	68.0	1786	7	US-10-750-185-31999	Sequence 31999, A
36	17	68.0	1786	7	US-10-750-623-31999	Sequence 31999, A
37	17	68.0	2038	8	US-11-054-281-29	Sequence 29, Appl
C 38	17	68.0	2494	7	US-10-750-185-27910	Sequence 27910, A
C 39	17	68.0	2494	7	US-10-750-623-27910	Sequence 27910, A
40	17	68.0	2583	7	US-10-947-243-67	Sequence 67, Appl
41	17	68.0	2716	7	US-10-995-561-19	Sequence 19, Appl
42	17	68.0	2735	7	US-10-750-185-25095	Sequence 25095, A
43	17	68.0	2735	7	US-10-750-623-25095	Sequence 25095, A
44	17	68.0	3824	7	US-10-750-185-29981	Sequence 29981, A
45	17	68.0	3824	7	US-10-750-623-29981	Sequence 29981, A
C 46	17	68.0	100000	8	US-11-124-368A-2898	Sequence 2898, Ap
47	17	68.0	403278	7	US-10-995-561-13421	Sequence 13421, A
48	17	68.0	1691140	8	US-11-091-018-1	Sequence 1, Appl
C 49	16.8	67.2	82	7	US-10-310-914A-4467	Sequence 4467, Ap
C 50	16.8	67.2	3152	7	US-10-793-626-4056	Sequence 4056, Ap
51	16.8	67.2	160213	8	US-11-121-086-103	Sequence 103, App
C 52	16.6	66.4	25	8	US-11-121-849-81092	Sequence 81092, A
53	16.6	66.4	25	8	US-11-136-527-350443	Sequence 350443, A
54	16.6	66.4	820	7	US-10-750-185-30816	Sequence 30816, A
55	16.6	66.4	820	7	US-10-750-623-30816	Sequence 30816, A
56	16.6	66.4	1400	8	US-11-136-527-8088	Sequence 8088, Ap
57	16.6	66.4	1475	8	US-11-136-527-3992	Sequence 3992, Ap
C 58	16.6	66.4	1789	7	US-10-750-185-64472	Sequence 64472, A
C 59	16.6	66.4	1789	7	US-10-750-623-64472	Sequence 64472, A
C 60	16.6	66.4	1953	7	US-10-750-185-48515	Sequence 48515, A
C 61	16.6	66.4	1953	7	US-10-750-623-48515	Sequence 48515, A
C 62	16.6	66.4	2764	7	US-10-750-185-58029	Sequence 58029, A
C 63	16.6	66.4	2764	7	US-10-750-623-58029	Sequence 58029, A
C 64	16.6	66.4	3660	8	US-11-136-527-3658	Sequence 3658, Ap
C 65	16.6	66.4	215308	7	US-11-121-086-77	Sequence 77, Appl
C 66	16.4	65.6	21	7	US-10-310-914A-963829	Sequence 963829, A
67	16.4	65.6	22	7	US-10-310-914A-498609	Sequence 498609, A
C 68	16.4	65.6	201	8	US-11-124-368A-14754	Sequence 14754, A
C 69	16.4	65.6	201	8	US-11-124-368A-14755	Sequence 14755, A
C 70	16.4	65.6	452	5	US-09-978-360A-318	Sequence 318, App
C 71	16.4	65.6	568	5	US-09-978-360A-387	Sequence 387, App
C 72	16.4	65.6	3065	7	US-10-750-185-40303	Sequence 40303, A
C 73	16.4	65.6	3065	7	US-10-750-623-40303	Sequence 40303, A
C 74	16.4	65.6	100000	8	US-11-124-368A-2313	Sequence 2313, Ap
C 75	16.4	65.6	199321	8	US-11-121-086-10	Sequence 10, Appl
C 76	16.2	64.8	25	8	US-11-136-527-350469	Sequence 350469, A
C 77	16.2	64.8	201	8	US-11-124-368A-14496	Sequence 14496, A
C 78	16.2	64.8	986	7	US-10-750-185-53782	Sequence 53782, A
C 79	16.2	64.8	986	7	US-10-750-623-53782	Sequence 53782, A
C 80	16.2	64.8	1141	7	US-10-750-185-36038	Sequence 36038, A
C 81	16.2	64.8	1141	7	US-10-750-623-36038	Sequence 36038, A
C 82	16.2	64.8	1690	8	US-11-090-439-43	Sequence 43, Appl
C 83	16.2	64.8	2666	7	US-10-821-234-215	Sequence 215, App
C 84	16.2	64.8	4162	8	US-11-136-527-2835	Sequence 2835, Ap
C 85	16.2	64.8	38527	8	US-11-124-368A-2312	Sequence 2312, Ap
C 86	16.2	64.8	154548	8	US-11-121-086-33	Sequence 33, Appl
C 87	16.2	64.8	168656	8	US-11-112-908-58	Sequence 58, Appl
C 88	16.2	64.8	170285	8	US-11-112-908-59	Sequence 59, Appl
C 89	16.2	64.8	191343	8	US-11-112-908-53	Sequence 53, Appl
C 90	16	64.0	201	7	US-10-995-561-32191	Sequence 32191, A
C 91	16	64.0	201	7	US-10-995-561-51277	Sequence 51277, A
C 92	16	64.0	600	8	US-11-136-527-4286	Sequence 4286, Ap
C 93	16	64.0	684	7	US-10-750-185-55780	Sequence 55780, A
C 94	16	64.0	684	7	US-10-750-623-55780	Sequence 55780, A
C 95	16	64.0	756	8	US-11-136-527-170	Sequence 170, App



```

96 16 64.0 762 8 US-11-037-243-39
c 97 16 64.0 796 7 US-10-750-185-60772
c 98 16 64.0 796 7 US-10-750-623-60772
c 99 16 64.0 962 7 US-10-750-185-52446
c 100 16 64.0 962 7 US-10-750-623-52446
c 101 16 64.0 970 7 US-10-750-185-49838
c 102 16 64.0 970 7 US-10-750-623-49838
c 103 16 64.0 1222 7 US-10-750-185-42634
c 104 16 64.0 1222 7 US-10-750-623-42634
c 105 16 64.0 1287 7 US-10-750-185-31665
c 106 16 64.0 1287 7 US-10-750-623-31665
c 107 16 64.0 1397 7 US-10-750-185-43816
c 108 16 64.0 1397 7 US-10-750-623-43816
c 109 16 64.0 1864 7 US-10-750-185-35675
c 110 16 64.0 1864 7 US-10-750-623-35675
c 111 16 64.0 1912 7 US-10-750-185-26763
c 112 16 64.0 1912 7 US-10-750-623-26763
c 113 16 64.0 2167 7 US-10-750-185-33805
c 114 16 64.0 2167 7 US-10-750-623-33805
c 115 16 64.0 2218 7 US-10-750-185-34556
c 116 16 64.0 2218 7 US-10-750-623-34556
c 117 16 64.0 2766 7 US-10-750-185-41252
c 118 16 64.0 2766 7 US-10-750-623-41252
c 119 16 64.0 3598 7 US-10-793-625-3488
c 120 16 64.0 4434 7 US-10-750-185-37523
c 121 16 64.0 4434 7 US-10-750-623-37523
c 122 16 64.0 10170 8 US-11-136-527-1817
c 123 16 64.0 35344 7 US-10-995-561-13307
c 124 16 64.0 100000 8 US-11-124-368A-2890
c 125 16 64.0 138821 8 US-11-121-086-80
c 126 16 64.0 150450 8 US-11-112-908-54
c 127 16 64.0 175023 8 US-11-121-086-18
c 128 16 64.0 179666 8 US-11-121-086-67
c 129 16 64.0 191343 8 US-11-112-908-53
c 130 16 64.0 387780 7 US-10-995-561-13259
c 131 15.8 63.2 19 9 US-11-101-244-824296
c 132 15.8 63.2 19 10 US-11-083-784-824296
c 133 15.8 63.2 24 7 US-10-310-914A-729831
c 134 15.8 63.2 25 7 US-10-750-185-23363
c 135 15.8 63.2 25 7 US-10-750-623-23363
c 136 15.8 63.2 25 8 US-11-121-849-603001
c 137 15.8 63.2 201 7 US-10-995-561-15042
c 138 15.8 63.2 600 7 US-10-750-185-4755
c 139 15.8 63.2 600 7 US-10-750-185-20853
c 140 15.8 63.2 600 7 US-10-750-623-4755
c 141 15.8 63.2 600 7 US-10-750-623-20853
c 142 15.8 63.2 1284 7 US-10-750-185-57551
c 143 15.8 63.2 1284 7 US-10-750-623-57551
c 144 15.8 63.2 1449 7 US-10-750-185-46171
c 145 15.8 63.2 1449 7 US-10-750-623-46171
c 146 15.8 63.2 2112 7 US-10-750-185-27053
c 147 15.8 63.2 2112 7 US-10-750-623-27053
c 148 15.8 63.2 2130 7 US-10-750-185-60713
c 149 15.8 63.2 2130 7 US-10-750-623-60713
c 150 15.8 63.2 2847 8 US-11-037-243-35

```

# ALIGNMENTS

```

RESULT 1
US-10-750-185-2224/c
; Sequence 2224, Application US/10750185
; Publication No. US200502603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

RESULT 3
US-11-136-527-3865
; Sequence 3865, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; APPLICANT: ROSENFELD, David
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3865

```

```

; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2224
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT12280
US-10-750-185-2224

Query Match 76.8%; Score 19.2; DB 7; Length 600;
Best Local Similarity 87.5%; Pred. No. 9,9;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACCTTGAAGGGAGACTCA 25
||||| ||||| ||||| ||||| ||
Db 397 AAAAACTCTTTGAAGGGAGACCA 374

RESULT 2
US-10-750-623-2224/c
; Sequence 2224, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2224
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT12280
US-10-750-623-2224

Query Match 76.8%; Score 19.2; DB 7; Length 600;
Best Local Similarity 87.5%; Pred. No. 9,9;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAACCTTGAAGGGAGACTCA 25
||||| ||||| ||||| ||||| ||
Db 397 AAAAACTCTTTGAAGGGAGACCA 374

```

; LENGTH: 777  
 ; TYPE: DNA  
 ; ORGANISM: R  
 US-11-136-527-3

Query Match 76.8%; Score 19.2; DB 8; Length 777;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

[illegible]

```

RESULT 4
US-11-136-527-7961
; Sequence 7961, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7961
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-7961

```

```

Query Match      76.8%; Score 19.2; DB 8; Length 777;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

**Qy**

2 AAAAAACACTTGGGAAGGGAAGCTCA 25  
||| ||| ||| ||| ||| ||| ||| |||  
**pB**

567 AAAGACACTTGCAAGGTGACTCA 590

```

RESULT 5
US-10-750-185-54511/c
; Sequence 54511, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54511
; LENGTH: 3221
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-54511

```

Query Match	75.8%	Score 19.2;	DB 7;	Length 3221;
Best Local Similarity	87.5%;	Pred. No. 14;		
Matches 21: Conservative	0: Mismatches	3: Indels	0: Gaps	0:

Qy 2 AAAACACTTGAAGGGAGACTCA 25  
||| ||| ||| ||| ||| ||| ||| |||  
db 1164 AAAAAGCTTTTGAAGGGAGACCCA 1141

```

RESULT 6
US-10-750-623-54511/c
; Sequence 54511, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC..
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54511
; LENGTH: 3221
; TYPE: DNA
; ORGANISM: Bovine 19866880959619
US-10-750-623-54511

```

Query Match 76.8%; Score 19.2; DB 7; Length 3221;  
Best Local Similarity 87.5%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAACACTTGGAAAGGAGACTCA 25  
|||  
db 1164 AAAAAACTCTTTGAAAGGAGACCCA 1141

```

RESULT 7
US-10-750-185-27280/c
; Sequence 27280, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27280
; LENGTH: 3481
; TYPE: DNA
; ORGANISM: Bovine 19866880695666
US-10-750-185-27280

```

Query Match	75.2%	Score 18.8;	DB 7;	Length 3481;
Best Local Similarity	90.9%;	Pred. No. 22;		
Matches 20: Conservative	0;	Mismatches 2;	Indels 0;	Caps 0;

QY  
1 AAAAAACACTTGGAAAGGAGAC 22





Db 646 AAAGACACTTGCACGGTGACTCA 669  
|||||

RESULT 17  
US-10-750-185-42140/c  
; Sequence 42140, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; APPLICANT: BATES, Stephen  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42140  
; LENGTH: 1582  
; TYPE: DNA  
; ORGANISM: Bovine 19866880823702  
US-10-750-185-42140

Query Match 70.4%; Score 17.6; DB 7; Length 1582;  
Best Local Similarity 83.3%; Pred. No. 67;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACAACCTTGAAGGGGAGACTC 24  
|||||

Db 1249 AAAAATCCATGTAGGGGAGACTC 1226  
|||||

RESULT 18  
US-10-750-623-42140/c  
; Sequence 42140, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42140  
; LENGTH: 1582  
; TYPE: DNA  
; ORGANISM: Bovine 19866880823702  
US-10-750-623-42140

Query Match 70.4%; Score 17.6; DB 7; Length 1582;  
Best Local Similarity 83.3%; Pred. No. 67;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACAACCTTGAAGGGGAGACTC 24  
|||||

Db 1249 AAAAATCCATGTAGGGGAGACTC 1226  
|||||

RESULT 19  
US-10-750-185-36667/c  
; Sequence 36667, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 36667  
; LENGTH: 1698  
; TYPE: DNA  
; ORGANISM: Bovine 19866880843352  
US-10-750-185-36667

Query Match 70.4%; Score 17.6; DB 7; Length 1698;  
Best Local Similarity 83.3%; Pred. No. 68;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACAACCTTGAAGGGGAGACTCA 25  
|||||

Db 1377 AAAAATTTGAAGGGGATTTCA 1354  
|||||

RESULT 20  
US-10-750-623-36667/c  
; Sequence 36667, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 36667  
; LENGTH: 1698  
; TYPE: DNA  
; ORGANISM: Bovine 19866880843352  
US-10-750-623-36667

Query Match 70.4%; Score 17.6; DB 7; Length 1698;  
Best Local Similarity 83.3%; Pred. No. 68;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACAACCTTGAAGGGGAGACTCA 25  
|||||

Db 1377 AAAAATTTGAAGGGGATTTCA 1354  
|||||

RESULT 21  
US-10-750-185-53076

; Sequence 53076, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 53076  
; LENGTH: 1805  
; TYPE: DNA  
; ORGANISM: Bovine 19866881384060  
US-10-750-185-53076

Query Match 68.8%; Score 17.2; DB 7; Length 1805;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAACACTTGGAGGAGACTC 24  
|||||  
DB 992 AAAACACTTAGAAAGAGACTC 1013

RESULT 22  
US-10-750-623-53076  
; Sequence 53076, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 53076  
; LENGTH: 1805  
; TYPE: DNA  
; ORGANISM: Bovine 19866881384060  
US-10-750-623-53076

Query Match 68.8%; Score 17.2; DB 7; Length 1805;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAACACTTGGAGGAGACTC 24  
|||||  
DB 992 AAAACACTTAGAAAGAGACTC 1013

RESULT 23  
US-10-750-185-35807  
; Sequence 35807, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 35807  
; LENGTH: 2655  
; TYPE: DNA  
; ORGANISM: Bovine 19866880701412  
US-10-750-185-35807

Query Match 68.8%; Score 17.2; DB 7; Length 2655;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGAGAC 22  
|||||  
DB 2257 AACAAACACTGGAAGTGAGAC 2278

RESULT 24  
US-10-750-623-35807  
; Sequence 35807, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 35807  
; LENGTH: 2655  
; TYPE: DNA  
; ORGANISM: Bovine 19866880701412  
US-10-750-623-35807

Query Match 68.8%; Score 17.2; DB 7; Length 2655;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACACTTGGAGGAGAC 22  
|||||  
DB 2257 AACAAACACTGGAAGTGAGAC 2278

RESULT 25  
US-11-121-086-104  
; Sequence 104, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES

```
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 104
; LENGTH: 186442
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-104

Query Match      68.8%; Score 17.2; DB 8; Length 186442;
Best Local Similarity 86.4%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACTCA 22
    ||||| ||||| ||||| ||||| |||||
Db 128241 AAAAAACACTTGGAGGGAGACTCA 128262

RESULT 26
US-10-995-561-1528
; Sequence 1528, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1528
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-1528

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
Db 132 AAAAAACACTTGGAGGGAGACTCA 156

RESULT 27
US-10-995-561-1530
; Sequence 1530, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1530
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-1530

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 83;

Qy 1 AAAAAACACTTGGAGGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
Db 132 AAAAAACACTTGGAGGGAGACTCA 156

RESULT 28
US-10-995-561-1538
; Sequence 1538, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1538
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-1538

Query Match      68.0%; Score 17; DB 7; Length 201;
Best Local Similarity 80.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
Db 168 AAAAAACACTTGGAGGGAGACTCA 192

RESULT 29
US-10-750-185-42937
; Sequence 42937, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERRE, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42937
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-42937

Query Match      68.0%; Score 17; DB 7; Length 1047;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAGGGAGACTCA 25
    ||||| ||||| ||||| ||||| |||||
Db 89 AAAAAACACTTGGAGGGAGACTCA 113

RESULT 30
US-10-750-623-42937
```





```
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5656
; LENGTH: 1698
; TYPE: DNA
; ORGANISM: Bovine 19866880612383
US-10-750-623-5656

Query Match      68.0%; Score 17; DB 7; Length 1698;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGGAGACTCA 25
   ||||| ||||| ||||| ||||| |||||
Db 783 AAAAAACACTTAGAAGAGTGCTGA 759

RESULT 35
US-10-750-185-31999
; Sequence 31999, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31999
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Bovine 19866880848133
US-10-750-185-31999

Query Match      68.0%; Score 17; DB 7; Length 1786;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGGAGACTCA 25
   ||||| ||||| ||||| ||||| |||||
Db 681 AATAACAGATGGAAGAGAGACTGA 705

RESULT 36
US-10-750-623-31999
; Sequence 31999, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
```

```
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31999
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Bovine 19866880848133
US-10-750-623-31999

Query Match      68.0%; Score 17; DB 7; Length 1786;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGGAGACTCA 25
   ||||| ||||| ||||| ||||| |||||
Db 681 AATAACAGATGGAAGAGAGACTGA 705

RESULT 37
US-11-054-281-29
; Sequence 29, Application US/11054281
; Publication No. US20060013813A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240CIP
; CURRENT APPLICATION NUMBER: US/11/054,281
; CURRENT FILING DATE: 2005-02-08
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 10/044,564
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 2038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-054-281-29

Query Match      68.0%; Score 17; DB 8; Length 2038;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGAAGGGGAGACTCA 25
   ||||| ||||| ||||| ||||| |||||
Db 1719 AAAAGATGCTTGTAAAGGGAGATTCA 1743

RESULT 38
US-10-750-185-27910/c
; Sequence 27910, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
```

```

; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27910
; LENGTH: 2494
; TYPE: DNA
; ORGANISM: Bovine
; US-10-750-185-27910

```

Query Match 68.0%; Score 17; DB 7; Length 2494;  
Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACACTTGGAAAGGGAGACTCA 25  
||| ||| ||| ||| ||| ||| ||| |||  
Db 1477 AAAAAAGCATTTTGAAGTGACACTCA 1453

RESULT 39  
US-10-750-623-27910/c  
; Sequence 27910, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: Denise, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27910  
; LENGTH: 2494  
; TYPE: DNA  
; ORGANISM: Bovine 19866880779532  
US-10-750-623-27910

Query Match	68.0%	Score 17	DB 7	Length 2494
Best Local Similarity	80.0%	Pred. No.	1.4e+02	
Matches	20	Conservative	0	Mismatches 5
		Indels	0	Gaps 0

QY 1 AAAAAACACTTGGAGGGAGACTCA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 1477 AAAAAGCATTTTGAAGTGACACTCA 1453

RESULT 40  
US-10-947-249-67/C  
; Sequence 67, Application US/10947249  
; Publication No. US20050287541A1  
; GENERAL INFORMATION:  
; APPLICANT: Akira NAKAGAWARA  
; APPLICANT: Miki OHIRA  
; APPLICANT: Shin ISHII  
; APPLICANT: Takeshi GOTO

```

; APPLICANT: HIROYUKI KUBO
; APPLICANT: TAKAHIRO HIRATA
; APPLICANT: YASUKO YOSHIDA
; APPLICANT: SAICHI YAMADA
; TITLE OF INVENTION: MICROARRAY FOR PREDICTING THE PROGNOSIS OF NEUROBLASTOMA
; TITLE OF INVENTION: PREDICTING THE PROGNOSIS OF NEUROBLASTOMA
; FILE REFERENCE: 117007
; CURRENT APPLICATION NUMBER: US/10/947,249
; CURRENT FILING DATE: 2004-09-23
; PRIOR APPLICATION NUMBER: US 60/505,614
; PRIOR APPLICATION NUMBER: 2003-09-25
; NUMBER OF SEQ ID NOS: 200
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 2583
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-947-249-67

```

Query Match	68.0%	Score 17;	DB 7;	Length 2583;
Best Local Similarity	80.0%;	Pred. No. 1.4e+02;		
Matches	20;	Conservative	0;	Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACACTTGGAAAGGGAGACTCA 25  
||||| ||||| ||||| |||||  
pb 1019 AAAAAAAGCTTGGGAATGCAGGCTCA 995

```

RESULT 41
US-10-995-561-19
; Sequence 19, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-19

```

Query Match	68.0%	Score 17	DB 7	Length 2716
Best Local Similarity	80.0%	Pred. No.	1.4e+02	
Matches	20	Mismatches	5	Indels 0
		Conservative		Gaps 0

Qy 1 AAAAAACACTTGGAGGGAGACTCA 25  
|||||  
2240 AAAAGATGCTTGTAAAGGGAGATTCA 2264  
Db

```

RESULT 42
US-10-750-185-25095
; Sequence 25095, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
;

```





QY 1 AAAAAACACTTCGAAGGGAG 20  
| | | | | | | | | | | | | | | |  
Db 1182 AAAAAACATTTCGAAGGGAG 1163

Search completed: February 3, 2006, 16:19:56  
Job time : 347.111 secs

:

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-15  
Perfect score: 25  
Sequence: 1 aaaaacacacacagcgagccctcg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_sts.\*  
11: gb\_sy.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	2331	9 BC005588	Mus muscu
2	23.4	93.6	17372	9 AC117831	AC117831 Mus muscu
3	23.4	93.6	205272	9 AC118646	AC118646 Mus muscu
4	23.4	93.6	246494	14 AC120454	AC120454 Rattus no
5	19.2	76.8	110000	2 CP000079_06	Continuation (7 of
6	19.2	76.8	127380	2 AC113585	AC113585 Trypanoso
7	18.8	75.2	2426	5 BC071044	BC071044 Xenopus 1
8	18.8	75.2	149754	9 AC158827	AC158827 Mus muscu
9	18.8	75.2	163916	14 AC148067	AC148067 Canis fam
10	18.6	74.4	791	5 CR386415	CR386415 Gallus ga
11	18.6	74.4	10029	1 AB013307	AB013307 Methanosa
12	18.6	74.4	54569	9 BX571685	BX571685 Mouse DNA
13	18.6	74.4	65758	14 AC100392	AC100392 Mus muscu
14	18.6	74.4	97626	9 AL845429	AL845429 Mouse DNA
15	18.6	74.4	109295	14 AC141049	AC141049 Rattus no
16	18.6	74.4	110000	14 AC119467_1	Continuation (2 of
17	18.6	74.4	110000	15 AP008216_225	Continuation (226
18	18.6	74.4	139999	15 AC018727	AC018727 Oryza sat

C	19	18.6	74.4	156342	9	AC158566	AC158566 Mus muscu
	20	18.6	74.4	157085	14	AC160549	AC160549 Mus muscu
	21	18.6	74.4	192561	9	AL805943	AL805943 Mouse DNA
	22	18.6	74.4	196097	8	AC122175	AC122175 Pan trogl
	23	18.6	74.4	214864	14	AC162627	AC162627 Bos tauru
	24	18.6	74.4	218219	14	AC154992	AC154992 Bos tauru
	25	18.6	74.4	218746	9	AC115721	AC115721 Mus muscu
	26	18.6	74.4	230451	14	AC138614	AC138614 Mus muscu
C	27	18.6	74.4	236746	14	AC119312	AC119312 Rattus no
	28	18.6	74.4	258567	14	CR376791	CR376791 Mus muscu
	29	18.6	74.4	300029	15	AB017122	AB017122 Oryza sat
C	30	18.4	73.6	142359	14	EX322621	EX322621 Danio rer
	31	18.4	73.6	158420	14	CR936300	CR936300 Danio rer
C	32	18.2	72.8	252	15	CUSCC	D63386 Cuscula sat
	33	18.2	72.8	1181	8	HS3232695	AJ323695 Homo sapi
	34	18.2	72.8	1824	13	AY374526	AY374526 Borna dis
	35	18.2	72.8	1824	13	AY374527	AY374527 Borna dis
C	36	18.2	72.8	10029	1	AE011319	AE011319 Leptospir
	37	18.2	72.8	37857	14	AC142512	AC142512 Rattus no
C	38	18.2	72.8	148875	9	AC101757	AC101757 Mus muscu
	39	18.2	72.8	150313	8	CR376791	AC137672 Homo sapi
C	40	18.2	72.8	156107	8	AC099543	AC099543 Homo sapi
	41	18.2	72.8	190122	14	AC083820	AC083820 Rattus no
C	42	18.2	72.8	212822	14	AC080156	AC080156 Rattus no
	43	18.2	72.8	225030	9	AC107608	AC107608 Rattus no
C	44	18.2	72.8	247137	14	AC098268	AC098268 Rattus no
	45	18.2	72.8	249741	14	AC096330	AC096330 Rattus no
C	46	18.2	72.8	271995	14	AC080155	AC080155 Rattus no
	47	18.2	72.8	301258	1	AE017296	AE017296 Leptospir
	48	18.2	72.8	349751	1	CR378680	CR378680 Photobact
C	49	17.8	71.2	1368	6	AX432478	AX432478 Sequence
	50	17.8	71.2	2347	6	AX954557	AX954557 Sequence
C	51	17.8	71.2	110000	1	AE017333_01	Continuation (2 of
	52	17.8	71.2	110000	1	CP000002_01	Continuation (2 of
C	53	17.8	71.2	115780	6	CQ869701_01	CQ869701 Sequence
	54	17.8	71.2	119077	14	AC147494	AC147494 Carolia
	55	17.8	71.2	178882	14	AC158270	AC158270 Carolia
	56	17.8	71.2	208684	9	AC122272	AC122272 Mus muscu
	57	17.8	71.2	215602	9	AC155330	AC155330 Mus muscu
	58	17.8	71.2	225535	14	AC111381	AC111381 Rattus no
	59	17.6	70.4	599	6	AX438410	AX438410 Sequence
	60	17.6	70.4	597	6	AX438347	AX438347 Sequence
	61	17.6	70.4	2018	15	MGU12335	U12335 Magnaporthe
	62	17.6	70.4	6073	6	CQ490795	CQ490795 Sequence
	63	17.6	70.4	6073	6	CQ496641	CQ496641 Sequence
C	64	17.6	70.4	53675	8	AL3391708	AL3391708 Human DNA
C	65	17.6	70.4	69058	14	AC165884	AC165884 Bos tauru
	66	17.6	70.4	69358	14	AC164502	AC164502 Bos tauru
C	67	17.6	70.4	94695	15	ATF23E13	AL022141 Arabidops
	68	17.6	70.4	110000	1	AE016822_05	Continuation (6 of
	69	17.6	70.4	110000	1	AY596297_06	Continuation (7 of
C	70	17.6	70.4	110000	1	AY596297_07	Continuation (8 of
	71	17.6	70.4	110000	1	CP000058_40	Continuation (41 o
	72	17.6	70.4	110000	14	AC073860_2	Continuation (3 of
C	73	17.6	70.4	110000	15	AP008212_080	Continuation (81 o
	74	17.6	70.4	114691	4	AC144636	AC144636 Artibeus
	75	17.6	70.4	114691	5	CR847979	CR847979 Zebrafish
	76	17.6	70.4	117036	8	AL136162	AL136162 Human DNA
	77	17.6	70.4	129135	14	AC023220	AC023220 Homo sapi
	78	17.6	70.4	145342	15	AP003490	AP003490 Oryza sat
	79	17.6	70.4	153734	15	AP005518	AP005518 Oryza sat
	80	17.6	70.4	155544	15	AP005518	AP005518 Oryza sat
	81	17.6	70.4	158431	14	AC147528	AC147528 Oryza sat
	82	17.6	70.4	161023	14	AC069508	AC069508 Homo sapi
C	83	17.6	70.4	172894	14	AC150571	AC150571 Bos tauru
	84	17.6	70.4	172963	7	AY962392	AY962392 Aeromonas
C	85	17.6	70.4	173591	7	AY375531	AY375531 Bacteriop
	86	17.6	70.4	177640	8	AC079456	AC079456 Homo sapi
C	87	17.6	70.4	178079	14	AC068929	AC068929 Homo sapi
	88	17.6	70.4	180223	8	AC062017	AC062017 Homo sapi
	89	17.6	70.4	180727	9	AC140278	AC140278 Mus muscu
	90	17.6	70.4	183008	14	AC019276	AC019276 Homo sapi
	91	17.6	70.4	188255	14	AC074241	AC074241 Homo sapi



## ORGANISM

## Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 2331)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bustow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheet, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2331)  
NIH MGC Project  
Direct Submission  
Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [angbcm.tmc.edu](mailto:angbcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loubege, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 17 Row: m Column: 23  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

## FEATURES

## source

1..2331  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:11579 IMAGE:3708463"  
/cissue\_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."  
/clone\_lib="NCI CGAP\_Mam6"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..2331  
/gene="Tegt"  
/db\_xref="GeneID:110213"  
/db\_xref="MGI:99682"  
68..781  
/gene="Tegt"

## gene

## CDS

AC096730 Homo sapi  
AL592287 Human DNA  
AC013624 Homo sapi  
AC021016 Homo sapi  
AL045282 Zebrafish  
AL096870 Human chr  
AL161589 Arabidops  
BX546480 Zebrafish  
AC146621 Ootolemur  
AC150988 Bos tauru  
AC094760 Rattus no  
AC087876 Bos tauru  
AC149658 Bos tauru  
AC156053 Bos tauru  
AL390960 Human DNA  
Continuation (19 o  
CR759824 Zebrafish  
AC106564 Rattus no  
BV597660 S216P6009  
BV573357 S215P6141  
AY500497 Peziza ar  
BC086711 Panio rer  
AY730404 Daphnia l  
AB097523 Macaca fa  
AC079961 Homo sapi  
AL031229 Human DNA  
AC106802 Homo sapi  
Continuation (16 o  
AL663040 Mouse DNA  
AL808126 Mouse DNA  
AL031768 Human DNA  
AC148527 Medicago  
Continuation (16 o  
Continuation (16 o  
AC109710 Rattus no  
Continuation (3 of  
Continuation (4 of  
Continuation (3 of  
Continuation (5 of  
Continuation (56 o  
Continuation (63 o  
Continuation (254  
Continuation (16 o  
Continuation (17 o  
AC012188 Sequence  
AC147012 Medicago  
AC022417 Homo sapi  
AP000559 Oryza sat  
AL060997 Oryza sat  
AC016410 Homo sapi  
AC130432 Homo sapi  
AC132214 Genomic s  
BX649302 Zebrafish  
AP000391 Oryza sat  
AL627362 DNA centr  
AC148232 Didelphis  
AL662969 Oryza sat  
AC105928 Oryza sat  
AC165385 Oryctolag

## ALIGNMENTS

RESULT 1  
BC005588 2331 bp mRNA linear ROD 25-JUL-2005  
LOCUS Mus musculus testis enhanced gene transcript, mRNA (cDNA clone  
DEFINITION MGC:11579 IMAGE:3708463), complete cds.  
ACCESSION BC005588  
VERSION BC005588.1 GI:13542767  
KEYWORDS Mus musculus (house mouse)  
SOURCE

```

/codon_start=1
/product="testis enhanced gene transcript"
/protein_id="AA0405588.1"
/db_xref="GI:13542768"
/db_xref="GeneID:110213"
/db_xref="MG1:99682"
/translations="MWIFDKINFDFALLKFSHTPTQOHLKKVYASFALCMFVAAAG
AYVHVTHRTFQAGLSALMIWLMATPHSHETQKRLGLLGAFLTGVLGCPA
LELCIAVNSILPTAPMGYAMITFCFSLSALYARRSYLFLGILMSANSLMLDLSLIG
NLFPGSIWLFQANLYIGLLVMCGFVFLDTQIIEKAHGDYKDIYHCVDLFLDFVTLF
RKLMLILAFNEKDKKKEK"

```

## ORIGIN

```

Query Match      93.6%  Score 23.4;  DB 9;  Length 2331;
Best Local Similarity 96.0%  Pred. No. 3.4;
Matches 24;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

```

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25

Db 2008 AAAAAACGACACTCGGAGCCCTCGG 2032

## RESULT 2

```

AC117831      173772 bp  DNA  linear  ROD 17-FEB-2005
Mus musculus chromosome 15, clone RP24-106E17, complete sequence.

```

## LOCUS

AC117831

## DEFINITION

AC117831

## ACCESSION

AC117831

## VERSION

AC117831.6

## KEYWORDS

HTG.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 173772)

Birren,B., Nussbaum,C. and Lander,E.

Mus musculus chromosome 15, clone RP24-106E17

Unpublished

2 (bases 1 to 173772)

Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,

Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,

Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faroo,S., Ferreira,P., FitzGerald,M., Gage,D., Gargan,J., Gardyna,S.,

Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,

Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,

Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,

Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,

Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollaxa,V.,

Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Traversman,A., Talamas,J., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 173772)

Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,

Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,

Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,

Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,

Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Traversman,A., Talamas,J., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.

## TITLE

JOURNAL

Submitted (20-JAN-2005) Broad Institute of MIT and Harvard, 320

Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 173772)

Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,

Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,

Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,

Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,

Erickson,J., Faroo,S., Ferreira,P., FitzGerald,M., Gage,D.,

Gargan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,

Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,

Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,

Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,

Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,

McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,

Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,

O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,

Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,

Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,

Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Traversman,A., Talamas,J., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,

Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (17-FEB-2005) Broad Institute of MIT and Harvard, 320

Charles Street, Cambridge, MA 02141, USA

On Feb 17, 2005 this sequence version replaced gi:57977528.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@broad.mit.edu

----- Project Information

Center project name: L25872

Center clone name: 106\_E\_17

-----

## FEATURES

source

1. 173772

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-106E17"

/clone\_lib="RPCI-24 Male Mouse BAC"

859..1005

/rpt\_family="B1\_MM"

1021..1197

/rpt\_family="B3"

1440..1550

/rpt\_family="RSINE1"

2148..2168

Some of the sequence contained within base pairs 119388 to the end  
of the clone was stolen from accession AC139317.

repeat\_region /rpt\_family="AT\_rich"  
2412..2435  
repeat\_region /rpt\_family="(TTTG)n"  
2811..2942  
repeat\_region /rpt\_family="B1\_MM"  
2943..2979  
repeat\_region /rpt\_family="(CA)n"  
complement(2985..3100)  
repeat\_region /rpt\_family="B2\_Mm2"  
complement(3186..3678)  
repeat\_region /rpt\_family="PB1D9"  
complement(3690..3894)  
repeat\_region /rpt\_family="MER82"  
4129..4187  
repeat\_region /rpt\_family="(TG)n"  
5087..5112  
repeat\_region /rpt\_family="AT\_rich"  
5203..5290  
repeat\_region /rpt\_family="B1\_MM"  
complement(6980..7144)  
repeat\_region /rpt\_family="ID\_B1"  
complement(8010..8148)  
repeat\_region /rpt\_family="B1\_MM"  
complement(8160..8341)  
repeat\_region /rpt\_family="B3"  
complement(10674..10867)  
repeat\_region /rpt\_family="ORR1D"  
complement(11222..11321)  
repeat\_region /rpt\_family="RSINE1"  
complement(12570..12717)  
repeat\_region /rpt\_family="B1\_MM"  
complement(12967..13089)  
repeat\_region /rpt\_family="B3"  
complement(13073..13201)  
repeat\_region /rpt\_family="ID\_B1"  
13265..13298  
repeat\_region /rpt\_family="(CG)n"  
complement(14043..14228)  
repeat\_region /rpt\_family="B2\_Mm1"  
14448..14641  
repeat\_region /rpt\_family="B3A"  
14654..14788  
repeat\_region /rpt\_family="B1\_MM"  
14790..14816  
repeat\_region /rpt\_family="(TAAAA)n"  
complement(14948..15137)  
repeat\_region /rpt\_family="B2\_Mm1"  
15188..15204  
repeat\_region /rpt\_family="AT\_rich"  
15205..15280  
repeat\_region /rpt\_family="ID4"  
15281..15311  
repeat\_region /rpt\_family="AT\_rich"  
15312..15430  
repeat\_region /rpt\_family="B1F"  
15427..15501  
repeat\_region /rpt\_family="ID\_B1"  
15559..15667  
repeat\_region /rpt\_family="(TA)n"  
15670..15783  
repeat\_region /rpt\_family="L1MC/D"  
complement(15799..15857)  
repeat\_region /rpt\_family="B3A"  
complement(15828..15857)  
repeat\_region /rpt\_family="B2\_Mm2"  
15858..15899  
repeat\_region /rpt\_family="(TTTTC)n"  
15903..15961  
repeat\_region /rpt\_family="(TC)n"  
complement(15962..16108)  
repeat\_region /rpt\_family="B2\_Mm2"  
complement(16129..16356)  
repeat\_region /rpt\_family="ID\_B1"

repeat\_region 17030..17068  
/rpt\_family="T-rich"  
repeat\_region 17132..17191  
/rpt\_family="PB1D7"  
repeat\_region 17192..17220

Query Match 93.6%; Score 23.4; DB 9; Length 173772;  
Best Local Similarity 96.0%; Pred. No. 2.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACACACAGCGAGCCCTCGG 25  
|||||  
Db 9971 AAAAAACGACACTGCGAGCCCTCGG 9995  
|||||

RESULT 3  
AC118646 205272 bp DNA linear ROD 25-DEC-2004  
Locus Mus musculus chromosome 15, clone RP24-116A18, complete sequence.

DEFINITION  
AC118646  
VERSION AC118646.23 GI:56799106  
KEYWORDS  
SOURCE HTG.  
ORGANISM Mus musculus (house mouse)

REFERENCE  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus chromosome 15, clone RP24-116A18  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 205272)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (19-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 205272)

REFERENCE  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Ericsson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,

Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemбек,L.,  
Zimmer,A. and Zody,M.

# Direct Submission

Submitted (06-NOV-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

## AUTHORS

4. (bases 1 to 205272)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavskiy,L., Bouhgaltier,B., Canarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArlano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kanat,A., Karatas,A., Kella,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,I., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemбек,L.,  
Zimmer,A. and Zody,M.

# Direct Submission

Submitted (25-DEC-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 25, 2004 this sequence version replaced gi:55468782.

All repeats were identified using RepeatMasker:

Smith, A.P.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information

Center project name: L25047

Center clone name: 116\_A\_18

-----

----- Location/Qualifiers

1. 205272

/organism="Mus musculus"

/mol\_type="Genomic DNA"

/db\_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-116A18"

/clone\_lib="RPCI-24 Male Mouse BAC"

1692. .1786

/rpt\_family="MIR3"

2376. .2415

/rpt\_family="MIR"

3053. 3097

/rpt\_family="(TG)n"

3369. .3443

/rpt\_family="CT-rich"

4222. .4368

/rpt\_family="RSINE1"

4552. 4732

/rpt\_family="C-rich"

5224. .5370

/rpt\_family="B1\_MM"

5372. 5393

/rpt\_family="(CAAA)n"

6288. .6314

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="(CAG)n"  
8243. .8271  
/rpt\_family="(TG)n"  
8714. .8745  
/rpt\_family="(T)n"  
8825. .8948  
/rpt\_family="PB1"  
8949. .8995  
/rpt\_family="GA-rich"  
9009. .9167  
/rpt\_family="ID\_B1"  
9168. .9292  
/rpt\_family="B1\_MM"  
9312. .9385  
/rpt\_family="ID\_B1"  
9402. .9423  
/rpt\_family="(TAAA)n"  
complement(9603. .9710)  
/rpt\_family="ID\_B1"  
complement(10551. .10700)  
/rpt\_family="B1\_MM"  
complement(10728. .10821)  
/rpt\_family="B1\_MM"  
11073. .11208  
/rpt\_family="B1\_MM"  
11209. .11257  
/rpt\_family="(GAAAA)n"  
complement(11917. .12096)  
/rpt\_family="B3"  
12128. .12275  
/rpt\_family="B1\_MM"  
complement(12537. .12824)  
/rpt\_family="ORR1D"  
complement(13530. .13623)  
/rpt\_family="PB1D7"  
13637. .13735  
/rpt\_family="(TCC)n"  
complement(13739. .13898)  
/rpt\_family="B3A"  
14887. .15008  
/rpt\_family="ID\_B1"  
16256. .16290  
/rpt\_family="AT rich"  
complement(18706. .18855)  
/rpt\_family="B1\_MM"  
19417. .19456  
/rpt\_family="(CAAAA)n"  
21051. .21124  
/rpt\_family="B1\_MM"  
21435. .21477  
/rpt\_family="(TCTCTG)n"  
complement(21484. .21592)  
/rpt\_family="B1\_MM"  
complement(22963. .23036)  
/notes="single clone coverage"  
24323. .24427  
/rpt\_family="B1F"  
26361. .26496  
/rpt\_family="MTD"  
26496. .26566  
/rpt\_family="MTD"  
26567. .26759  
/rpt\_family="B2\_Mm1"  
26760. .26847  
/rpt\_family="MTD"  
26931. .26953  
/rpt\_family="(TTTG)n"  
complement(27072. .27412)  
/rpt\_family="MTE"  
complement(27529. .27656)  
/rpt\_family="B1\_MM"

Query Match

93.6%; Score 23.4; DB 9; Length 205272;

Best Local Similarity 96.0%; Pred. No. 2,7; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 183211 AAAAAACGACACTGGGAGCCCTCGG 183235

RESULT 4  
AC120454 246494 bp DNA linear HTG 15-NOV-2002  
LOCUS Rattus norvegicus clone CH230-272K4, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 15 unordered pieces.  
AC120454  
AC120454.5 GI:25008669  
VERSION HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.  
1 (bases 1 to 246494)

REFERENCE  
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,C.A., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hludyn,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhera,L., Louleghed,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,F., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished

TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

2 (bases 1 to 246494)  
Worley,K.C.  
Direct Submission  
Submitted (07-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 246494)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23611251.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWKZ  
Center clone name: CH230-272K4  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 202515 bases at least Q40  
Consensus quality: 207572 bases at least Q30  
Consensus quality: 211369 bases at least Q20  
Estimated insert size: 205757; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 223610: contig of 223610 bp in length  
\* 223611 223710: gap of unknown length  
\* 223711 225139: contig of 1429 bp in length  
\* 225140 225239: gap of unknown length  
\* 225240 226743: contig of 1504 bp in length  
\* 226744 226843: gap of unknown length  
\* 226844 228228: contig of 1285 bp in length  
\* 228129 228228: gap of unknown length  
\* 228229 229753: contig of 1525 bp in length  
\* 229754 229853: gap of unknown length  
\* 229854 231650: contig of 1797 bp in length  
\* 231651 231750: gap of unknown length  
\* 231751 233063: contig of 1313 bp in length  
\* 233064 233163: gap of unknown length  
\* 233164 234374: gap of unknown length  
\* 234375 235660: contig of 1286 bp in length  
\* 235661 235760: gap of unknown length  
\* 235761 237660: contig of 1900 bp in length  
\* 237661 239922: contig of 2162 bp in length  
\* 237761 240022: gap of unknown length  
\* 239923

\* 240023 241516: contig of 1494 bp in length  
 \* 241517 241616: gap of unknown length  
 \* 241617 242741: contig of 1125 bp in length  
 \* 242742 242841: gap of unknown length  
 \* 242842 244151: contig of 1310 bp in length  
 \* 244152 244251: gap of unknown length  
 \* 244252 246494: contig of 2243 bp in length.

## FEATURES

source 1..246494  
 Location/Qualifiers  
 1..246494  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-272K4"  
 620..1244  
 /note="clone boundary"  
 misc\_feature clone\_end:Sp6  
 site:  
 end sequence:BZ240024"  
 12216..13388  
 /note="wgs contig"  
 misc\_feature complement(220115..220908)  
 misc\_feature /note="clone boundary"  
 clone\_end:T7  
 site:  
 end sequence:BZ240021"  
 223611..223710  
 /estimated\_length=unknown  
 gap 225140..225239  
 /estimated\_length=unknown  
 gap 226744..226843  
 /estimated\_length=unknown  
 gap 228129..228228  
 /estimated\_length=unknown  
 gap 229754..229853  
 /estimated\_length=unknown  
 gap 231651..231750  
 /estimated\_length=unknown  
 gap 233064..233163  
 /estimated\_length=unknown  
 gap 234275..234374  
 /estimated\_length=unknown  
 gap 235661..235760  
 /estimated\_length=unknown  
 gap 237661..237760  
 /estimated\_length=unknown  
 gap 239923..240022  
 /estimated\_length=unknown  
 gap 241517..241616  
 /estimated\_length=unknown  
 gap 242742..242841  
 /estimated\_length=unknown  
 gap 244152..244251  
 /estimated\_length=unknown

## ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 246494;  
 Best Local Similarity 96.0%; Pred. No. 2.7;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGGCCCTCG 25  
 DB 231986 AAAAAACGACACGCGGCCCTCG 232010

RESULT 5  
 CP000079\_06/c

## WPCOMMENT

Sequence split into 12 fragments LOCUS CP000079 Accession CP000079  
 Fragment Name Begin End  
 CP000079\_00 1 110000  
 CP000079\_01 100001 210000  
 CP000079\_02 200001 310000  
 CP000079\_03 300001 410000

CP000079\_04 400001 510000  
 CP000079\_05 500001 610000  
 CP000079\_06 600001 710000  
 CP000079\_07 700001 810000  
 CP000079\_08 800001 910000  
 CP000079\_09 900001 1010000  
 CP000079\_10 1000001 1110000  
 CP000079\_11 1100001 1130447

Continuation (7 of 12) of CP000079 from base 600001 (CP000079 Leishmania major strain F

Query Match 76.8%; Score 19.2; DB 2; Length 110000;  
 Best Local Similarity 87.5%; Pred. No. 2.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGGCCCTCG 24  
 DB 14660 AAAAAACGACACGCGGCCCTCG 14637

## RESULT 6

AC113585

## LOCUS

DEFINITION

AC113585

Trypanosoma brucei chromosome 5 clone RPI93-27M3, complete

sequence.

AC113585

AC113585.4 GI:62359100

HTG.

Trypanosoma brucei

Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE

1 (bases 1 to 127380)

Ghedini, E., Blandin, G., Bartholomeu, D., Caler, E., Haas, B.,

Hannick, L., Shallom, J., Hou, L., Djikeng, A., Felblyum, T.,

Hostettler, J., Johnson, J., Jones, K., Koo, H. L., Larkin, C., Bai, G.,

Peterson, J., Khalak, H. G., Salzberg, S., Simpson, A. J., Tallon, L., Van

Aken, S., Wanless, D., White, O., Wortman, J., Fraser, C. M. and

El-Sayed, N. M. A.

Unpublished

2 (bases 1 to 127380)

El-Sayed, N. M., Khalak, H. and Adams, M. D.

Direct Submission

Submitted (02-MAR-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

3 (bases 1 to 127380)

El-Sayed, N. M., Khalak, H. and Adams, M. D.

Direct Submission

Submitted (05-APR-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

4 (bases 1 to 127380)

Haas, B., Blandin, G. and El-Sayed, N.

Direct Submission

Submitted (07-APR-2005) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, nelsayed@igr.org

On Apr 7, 2005 this sequence version replaced gi:19774505.

BAC clone RPI93-27M3 from Trypanosoma brucei chromosome 5.

Putative protein coding regions (>30 codons) were predicted by

combining the output of the gene finding algorithm GLIMMER with

similarity data and manual annotation of open reading frames.

Further refinement of gene models (additions, deletions and

alterations to boundary coordinates) were done manually based on

detailed comparisons with the T. cruzi and L. major genomes. Genes

on BACs were assigned a systematic name based on the chromosome and

the BAC from which they originated (e.g. Tb03.27F10.410). Gene

products were assigned a putative function when they shared

significant similarity with experimentally characterized gene

products or when they contained functionally known protein domains.

Gene products were labelled either 'hypothetical protein', or

'hypothetical protein, conserved' in the case of products showing

significant similarity with proteins of unknown function

from other organisms. Gene products of unknown function with

predicted orthologs in Trypanosoma cruzi and Leishmania major

genomes were automatically classified as conserved proteins. Short



predicted coding sequences (less than 150 codons) with no additional evidence of coding potential were labelled 'hypothetical protein, unlikely'.

```
FEATURES
    source
        1. .127380
            /organism="Trypanosoma brucei"
            /mol_type="genomic DNA"
            /isolate="GUTat10.1"
            /db_xref="taxon:5691"
            /chromosome="5"
            /clone="RPCI93-27M3"
        1. .672
            /note="annotation for this sequence region can be found on
            overlapping clone 26K5."
        <1078. .>2694
            /locus_tag="Tb927.5.3070"
            /note="synonym: Tb05.27M3.20; hypothetical protein,
            conserved"
        <1078. .>2694
            /locus_tag="Tb927.5.3070"
            /product="hypothetical protein, conserved"
        1078. .2694
            /locus_tag="Tb927.5.3070"
            /note="go function: zinc ion binding [goid 0008270]"
            /codon_start=1
            /product="hypothetical protein, conserved"
            /protein_id="AAK79547.1"
            /db_xref="GI:62359101"
            /translation="MAKKKSKRGVEVKDISNFOYGYRADSSNAVTPVATRRVPVDP
            AVLKRAYILHNFQVILRHDAQOQLRAVPGLPQRPQHVOTAEALRLSGTTPWMDLIIVAVR
            GTWEERPCVLEEPOAARTSCGHVCLVCLOYISRRKASKORTCPVCSNFTIVA
            SLRCPNVLVEPPSVGARASFTMLKRGKSCILLQDDPCWKETTLTDDDELRLPFVE
            SPATSYRISTEGATREDECTAIRQISMLEAKQRPFTSFDDLLKRAEDAIRL
            TLQGVPPQASGETSSPLPAKPTGTETNMAYEFYGTGGQAYNHPHYKMLCVDA
            EARNPLRVAVPVEEITFTQDEASRXYRFARVPLHATIKLCLDLSDIVLAST
            MKIFEPFLARLKRSLRESSISSSTEDSTQOYLRYRTDWRGEARTLEDMSPSE
            FMWSEMSADFERLPALELSPSGTPELPVPQASAGRETGGCGWKGTGNSLQLFSPPP
            VAPVPTWGHGFENKPNKSSNSVLABGRK"
        2852. .2879
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="ta"
            <3399. .>4403
            /locus_tag="Tb927.5.3080"
            /notes="synonym: Tb05.27M3.40; hypothetical protein,
            conserved"
        <3399. .>4403
            /locus_tag="Tb927.5.3080"
            /product="hypothetical protein, conserved"
        3399. .4403
            /locus_tag="Tb927.5.3080"
            /codon_start=1
            /product="hypothetical protein, conserved"
            /protein_id="AAK79548.1"
            /db_xref="GI:62359102"
            /translation="MWGPQEVNAHPHDCGKSYFLPSTSAAGAAALCASLLPHSNLVCA
            PLNPHIAREPVTATGGDGTQVALLTRFRIENERLDDQVKAHVQVQEAELTQLQELS
            ARNTALHQDFRRKNDNAVVEKLKKQLEVMEVRAQLQVGLNGRISAYBEAISR
            DHPQGLSANDRVRELI CEALSSAGFLTRVANAVEHCHPSTHEVTDCPRSVEECTNH
            CQAAMRGKELPGETAARGHIVSCIELASALREEVLFCTEAVRVAALKPFMERSAD
            SSNAADRGTGLQHEESAPLARPTAECAEAHEL PQSDRRPRRPAARPTVEDCEVQ
            "
        <4874. .>5179
            /locus_tag="Tb927.5.3090"
            /notes="synonym: Tb05.27M3.50; hypothetical protein,
            conserved"
        <4874. .>5179
            /locus_tag="Tb927.5.3090"
            /product="hypothetical protein, conserved"
        4874. .5179
            /locus_tag="Tb927.5.3090"
            /codon_start=1
            /product="hypothetical protein, conserved"
            /repeat_region
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="ta"
            12205. .12252
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="ta"
            complement(<12339. .>13280)
            /locus_tag="Tb927.5.3120"
            /note="synonym: Tb05.27M3.120; translation initiation
            factor, putative"
            complement(<12339. .>13280)
            /locus_tag="Tb927.5.3120"
            /product="translation initiation factor, putative"
            complement(12339. .13280)
            /locus_tag="Tb927.5.3120"
            /note="go component: cytoplasm [goid 0005737];
            go_function: translation initiation factor activity [goid
            0003743];
            go_process: translational initiation [goid 0006413]"
            /codon_start=1
```

```
repeat_region
    /protein_id="AAK79549.1"
    /db_xref="GI:62359103"
    /translation="WSSGFHFHVDNSDAIKGMPSEALHKHLENAQLAHRICLAKALK
    AGEPPVEKCALTWGEYLIRYQAWSEYRPFQDSVAQYKYKWSKKRQEDDKNPKF"
    8385. .8409
        /note="microsatellite"
        /rpt_type=tandem
        /rpt_unit="tg"
    9246. .9265
        /note="microsatellite"
        /rpt_type=tandem
        /rpt_unit="ta"
        complement(<9330. .>10094)
        /locus_tag="Tb927.5.3100"
        /note="synonym: Tb05.27M3.80; hypothetical protein"
        complement(<9330. .>10094)
        /locus_tag="Tb927.5.3100"
        /product="hypothetical protein"
        complement(9330. .10094)
        /locus_tag="Tb927.5.3100"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AAK79550.1"
        /db_xref="GI:62359104"
        /translation="MINKENGRVHYSTDVSSVGSKQGITHGQIGKSGRYNINFAIAW
        LDETMLDDPKLYTNHSTVEDEOGRMWSLETETTYQFITILLQYYPWVGRNVGADE
        VRRKIIDDLREALSRDITIKDLEAGSOLFRLIFQYENDWIRELGFSAGOLKRRHLT
        EKMIYVGLKLSALYIMNVNKGKQKNSCDGILRLREELSRERSGVNMLPPANPENSFP
        PPSVNGHPQTLPRALRNDELQAFATVTPDFEIRGE"
        10667. .10686
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="gaaaa"
            complement(<10702. .>11358)
            /locus_tag="Tb927.5.3110"
            /note="synonym: Tb05.27M3.100; hypothetical protein,
            conserved"
            complement(<10702. .>11358)
            /locus_tag="Tb927.5.3110"
            /product="hypothetical protein, conserved"
            /codon_start=1
            /product="hypothetical protein, conserved"
            /protein_id="AAK79551.1"
            /db_xref="GI:62359105"
            /translation="MLNRSRURRVWLGTAANTSIGRDVMPSTGKYNIVSPTDVAAN
            ISAGAMPEANVAALPVRIQYSKNRALSVDYDELLKLPLRYKLHDFAKLCPQQR
            COEPEAVGVVGRDSAVGYLPPLGPADPLDTIPFFVHRSSNGVLPGKFSMHRNIMP
            APYLRIQQVEGDMRFEPELLKIPTTKKIFVRSHSIYVYNVALDGRMVLHWWLLGLGF
            "
        11620. .11647
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="ta"
        12205. .12252
            /note="microsatellite"
            /rpt_type=tandem
            /rpt_unit="ta"
            complement(<12339. .>13280)
            /locus_tag="Tb927.5.3120"
            /note="synonym: Tb05.27M3.120; translation initiation
            factor, putative"
            complement(<12339. .>13280)
            /locus_tag="Tb927.5.3120"
            /product="translation initiation factor, putative"
            complement(12339. .13280)
            /locus_tag="Tb927.5.3120"
            /note="go component: cytoplasm [goid 0005737];
            go_function: translation initiation factor activity [goid
            0003743];
            go_process: translational initiation [goid 0006413]"
            /codon_start=1
```



/product="translation initiation factor, putative"  
/protein\_id="AAK79552.1"

Query Match 76.8%; Score 19.2; DB 2; Length 127380;

Best Local Similarity 87.5%; Pred. No. 2.9e+02; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCG 24

|||||  
Db 94165 AAAAAACGACGCGAACCCCTCG 94188

#### RESULT 7

##### BC071044

##### LOCUS

DEFINITION Xenopus laevis hypothetical protein MGC83180, mRNA (cDNA clone MGC:83180 IMAGE:6631780), complete cds.

##### ACCESSION

##### BC071044

##### VERSION

##### BC071044.1

##### GI:47506955

##### KEYWORDS

##### MGC.

##### SOURCE

##### ORGANISM

##### Xenopus laevis

##### (African clawed frog)

##### Xenopus laevis

##### Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

##### Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

##### Xenopodinae; Xenopus; Xenopus.

##### 1 (bases 1 to 2426)

##### Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.

##### and Richardson, P.

##### Genetic and genomic tools for Xenopus research: The NIH Xenopus

##### Initiative

##### Dev. Dyn. 225 (4), 384-391 (2002)

##### 12454917

##### 2 (bases 1 to 2426)

##### Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

##### Stausberg, R.D., Collins, F.S., Wagner, L., Shermen, C.M., Schuler, G.D.,

##### Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,

##### McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

##### Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

##### Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

##### Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,

##### Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

##### Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

##### Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

##### Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalius, D.E.,

##### Schmerer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

##### Generation and initial analysis of more than 15,000 full-length

##### human and mouse cDNA sequences

##### Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

##### 12477932

##### 3 (bases 1 to 2426)

##### Klein, S. and Strausberg, R.

##### Direct Submission

##### Submitted (18-MAY-2004) National Institutes of Health, Xenopus Gene

##### Collection (XGC), National Institute of Child Health and Human

##### Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD

##### 20892-7510, USA

##### NIH-MGC Project

##### Contact: XGC help desk

##### Email: cga@nsl.nih.gov

##### Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

##### CNA Library Preparation: Life Technologies, Inc.

##### CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

##### cDNA Sequencing by: Genome Sequence Centre,

##### BC Cancer Agency, Vancouver, BC, Canada

##### info@bcgsc.bc.ca

##### Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Cloutier, Ruth

##### Andy Chan, Steve S. Chand, William Chow, Alison Butterfield,

#### FEATURES

##### source

##### 1..2426

##### /organism="Xenopus laevis"

##### /mol\_type="mRNA"

##### /db\_xref="taxon:8355"

##### /clone="MGC:83180 IMAGE:6631780"

##### /tissue\_type="Embryo, stage 10"

##### /clone\_lib="NICHD\_XGC\_Emb1"

##### /lab\_host="DH10B"

##### /notes="Vector: PCMV-SPORT6"

##### 1..2426

##### /gene="MGC83180"

##### /db\_xref="GeneID:432098"

##### 89..946

##### /gene="MGC83180"

##### /codon\_start=1

##### /product="MGC83180 protein"

##### /protein\_id="AAH71044.1"

##### /db\_xref="GI:47506956"

##### /db\_xref="GeneID:432098"

##### /translation="MREVLQHQDPGTVEPPGLGRLEAHRIIVISLGIYHSRVQ

##### RGGILKHNLMVSLVRSAGQVYLSQSPLEQYVLRQALHNPQDCKEPCPL

##### TGCTPQPTQDADATCTSCPTASRGNPDNTENPRGVCQGHVTCRSHSLHPE

##### DICPSHEAPSPSPSSCCRRKSGNPDGPMKRAKREQGEVPEFGDEAEHME

##### TENVANLISIFGTFSGLSKPEPTAEAEPAQDVQELDSTGAGPGCSERAI

##### SNSTAIEAF"

##### ORIGIN

##### Query Match 75.2%; Score 18.8; DB 5; Length 2426;

##### Best Local Similarity 90.9%; Pred. No. 5.5e+02;

##### Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

##### QY 4 AAACGACACGAGCGCCCTCGG 25

##### |||||

##### Db 478 AAACGACACGAGACCCCTCGG 499

##### |||||

##### RESULT 8

##### AC155827/c

##### LOCUS

##### DEFINITION

##### Mus musculus 10 BAC RP24-408N5 (Roswell Park Cancer Institute

##### (C57BL/6J Male) Mouse BAC Library) complete sequence.

##### AC155827 AC116864

##### AC155827.10 GI:67972468

##### HTG.

##### Mus musculus (house mouse)

##### Mus musculus

##### Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

##### Mammalia; Euthera; Euarchontoglires; Glires; Rodentia;

##### Sciurognathi; Muridae; Murinae; Mus.

##### 1 (bases 1 to 149754)

##### Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P.,

##### Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,

##### Baraducci, A., Biswalo, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,

##### Cadore, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,

##### Chacko, J., Chabroux, M., Chavez, D., Chen, A., Chen, G., Chen, R.,

##### Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,

##### Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De

##### Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,

##### Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,

Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lilla Prabhau, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalls, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 157 Row: a Column: 21  
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

#### Location/Qualifiers

##### 1..2426

##### /organism="Xenopus laevis"

##### /mol\_type="mRNA"

##### /db\_xref="taxon:8355"

##### /clone="MGC:83180 IMAGE:6631780"

##### /tissue\_type="Embryo, stage 10"

##### /clone\_lib="NICHD\_XGC\_Emb1"

##### /lab\_host="DH10B"

##### /notes="Vector: PCMV-SPORT6"

##### 1..2426

##### /gene="MGC83180"

##### /db\_xref="GeneID:432098"

##### 89..946

##### /gene="MGC83180"

##### /codon\_start=1

##### /product="MGC83180 protein"

##### /protein\_id="AAH71044.1"

##### /db\_xref="GI:47506956"

##### /db\_xref="GeneID:432098"

##### /translation="MREVLQHQDPGTVEPPGLGRLEAHRIIVISLGIYHSRVQ

##### RGGILKHNLMVSLVRSAGQVYLSQSPLEQYVLRQALHNPQDCKEPCPL

##### TGCTPQPTQDADATCTSCPTASRGNPDNTENPRGVCQGHVTCRSHSLHPE

##### DICPSHEAPSPSPSSCCRRKSGNPDGPMKRAKREQGEVPEFGDEAEHME

##### TENVANLISIFGTFSGLSKPEPTAEAEPAQDVQELDSTGAGPGCSERAI

##### SNSTAIEAF"

Query Match 75.2%; Score 18.8; DB 5; Length 2426;

Best Local Similarity 90.9%; Pred. No. 5.5e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACGACACGAGCGCCCTCGG 25

|||||

Db 478 AAACGACACGAGACCCCTCGG 499

|||||

#### RESULT 8

##### AC155827/c

##### LOCUS

##### DEFINITION

##### Mus musculus 10 BAC RP24-408N5 (Roswell Park Cancer Institute

##### (C57BL/6J Male) Mouse BAC Library) complete sequence.

##### AC155827 AC116864

##### AC155827.10 GI:67972468

##### HTG.

##### Mus musculus (house mouse)

##### Mus musculus

##### Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

##### Mammalia; Euthera; Euarchontoglires; Glires; Rodentia;

##### Sciurognathi; Muridae; Murinae; Mus.

##### 1 (bases 1 to 149754)

##### Muzny, D., Adams, C., Agbai II, O., Allen, C., Alsbrooks, S., Archer, P.,

##### Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,

##### Baraducci, A., Biswalo, K., Blyth, P., Bonham, H., Buhay, C., Burch, P.,

##### Cadore, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,

##### Chacko, J., Chabroux, M., Chavez, D., Chen, A., Chen, G., Chen, R.,

##### Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,

##### Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De

##### Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,

##### Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,

Dziuda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P., Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T., Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haeblerlein, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B., Howell, L. T., Hulyk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jang, H., Johnson, B., Johnson, R., Kalafus, K., Kelly, S., Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F., Leal, S., Lee, K., Lee, S., Legall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linnell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorenzshew, L., Lozado, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McClelland, H., McPherson, J., Mercadao, C., Metzker, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M., Muray, D., Nazarith, L., Ngo, D., Nguyen, N., Norwig-Estaugh, E., Nott, A., Nwaokwelu, O., Obregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rives, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Saverly, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Usmani, K., Vargo, C., Verduzco, D., Villasana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 149754)  
Worley, K.C.  
Direct Submission  
Submitted (20-JAN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 149754)  
Worley, K.C.  
Direct Submission  
Submitted (01-MAR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 149754)  
Worley, K.C.  
Direct Submission  
Submitted (11-MAR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 149754)  
Worley, K.C.  
Direct Submission  
Submitted (18-JUN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 18, 2005 this sequence version replaced gi:60735107.  
Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

repeat\_region /rpt family="AT rich"  
complement(12248..12353)  
repeat\_region /rpt family="PBID10"  
13905..14043  
repeat\_region /rpt family="B2\_Mm2"  
14951..15017  
repeat\_region /rpt family="TA)n"  
15782..15814  
repeat\_region /rpt family="AT rich"  
complement(16006..16145)  
repeat\_region /rpt family="Blf"  
17478..17501  
repeat\_region /rpt family="AT rich"  
18588..18644  
repeat\_region /rpt family="(TCCA)n"  
complement(18653..18795)  
repeat\_region /rpt family="RSINE1"  
19137..19223  
repeat\_region /rpt family="MIR"  
19348..19406  
repeat\_region /rpt family="(TG)n"  
19576..19725  
repeat\_region /rpt family="Bl\_Mur3"  
19726..19762  
repeat\_region /rpt family="GA-rich"  
20018..20083  
repeat\_region /rpt family="PBID10"  
complement(20257..20326)  
repeat\_region /rpt family="ID4"  
20793..20819  
repeat\_region /rpt family="(CAAAA)n"  
21389..21428  
repeat\_region /rpt family="A-rich"

Query Match 75.2%; Score 18.8; DB 9; Length 149754;  
Best Local Similarity 90.9%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAACGACACGCGAGCCCTC 23  
||||| ||||| ||||| ||||| |||||  
Db 80249 AAAAACGACACGCGAGCCCTC 80228

RESULT 9  
AC148067  
LOCUS AC148067.2 163916 bp DNA linear HTG 15-APR-2004  
DEFINITION Canis familiaris clone RP81-30P2, WORKING DRAFT SEQUENCE, 9 ordered  
pieces.  
ACCESSION AC148067  
VERSION AC148067.2 GI:46391174  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
Canis.

REFERENCE 1 (bases 1 to 163916)  
AUTHORS Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,  
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,  
Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J.,  
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B.,  
Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Lari,P.,  
Larson,S., Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,  
Margulies,E.H., Masello,C., Maskeri,B., McDowell,J.,  
Mullikin,J.C., Paquirigan,C., Portnoy,M.E., Prasad,A., Puri,O.,  
Reddi-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C.,  
Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L.,  
Wetherby,K.D., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 163916)  
REFERENCE Green,E.D.  
AUTHORS Direct Submission  
TITLE

JOURNAL Submitted (29-JAN-2004) NIH Intramural Sequencing Center, 8717  
REFERENCE Groveomont Circle, Gaithersburg, MD 20877, USA  
AUTHORS 3 (bases 1 to 163916)  
TITLES Green,E.D.  
JOURNAL Direct Submission  
COMMENT Submitted (15-APR-2004) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
On Apr 15, 2004 this sequence version replaced gi:41386836.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nih.gov  
----- Project Information  
Center project name: flyg  
Center clone name: 030P02

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.950319  
Consensus quality: 162694 bases at least Q40  
Consensus quality: 162942 bases at least Q30  
Consensus quality: 163053 bases at least Q20  
Insert size: 149000; agarose-fp  
Insert size: 163116; sum-of-contigs  
Quality coverage: 11.40x in Q20 bases; agarose-fp  
Quality coverage: 10.42x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 10164: contig of 10164 bp in length  
\* 10165 10264: gap of unknown length  
\* 10265 13805: contig of 3541 bp in length  
\* 13806 13905: gap of unknown length  
\* 13906 16621: contig of 2716 bp in length  
\* 16622 16721: gap of unknown length  
\* 16722 60680: contig of 43959 bp in length  
\* 60681 60780: gap of unknown length  
\* 60781 78425: contig of 17645 bp in length  
\* 78426 78526: gap of unknown length  
\* 78527 100161: contig of 21636 bp in length  
\* 100162 100261: gap of unknown length  
\* 100262 103084: contig of 2823 bp in length  
\* 103085 103184: gap of unknown length  
\* 103185 105651: contig of 2467 bp in length  
\* 105652 105751: gap of unknown length  
\* 105752 163916: contig of 58165 bp in length.

FEATURES  
Location/Qualifiers  
1..163916  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9615"  
/clone="RP81-30P2"

sequencing project.  
 This sequence is from the  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
 from a library constructed by Elizabeth Bosch. cDNA was prepared  
 from RNA extracted from chondrocytes isolated from growth plates,  
 normalised,  
 and poly A-trimmed. EcoRI-NotI cut cDNA was then ligated into the  
 vector. Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI  
 Host: Escherichia coli DH10B.  
 Location/Qualifiers  
 1..791  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer and broiler"  
 /db\_xref="taxon:9031"  
 /clone="CHEST231n18"  
 /clone\_lib="CSEORBN09"  
 /dev\_stage="juvenile"

**FEATURES**  
 source  
 1..791  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer and broiler"  
 /db\_xref="taxon:9031"  
 /clone="CHEST231n18"  
 /clone\_lib="CSEORBN09"  
 /dev\_stage="juvenile"

**ORIGIN**  
 Query Match 74.4%; Score 18.6; DB 5; Length 791;  
 Best Local Similarity 84.0%; Pred. No. 7.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAACGACAGCGAGCCCTCGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 726 AAAAAATGACAAAGCGAGCTTCTG 702  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 11  
 AE013307 10029 bp DNA linear BCT 07-FEB-2003  
 LOCUS Methanosarcina mazei strain Goel, section 89 of the complete  
 DEFINITION genome.  
 ACCESSION AE013307 AE008384  
 VERSION AE013307.1 GI:20905228  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Methanosarcina mazei Goel  
 Methanosarcina mazei Goel  
 Archaea: Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 Methanosarcinaceae; Methanosarcina.  
 1 (bases 1 to 10029)  
 Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,  
 Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,  
 Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,  
 Bhattacharya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,  
 Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.  
 The genome of Methanosarcina mazei: evidence for lateral gene  
 transfer between bacteria and archaea  
 J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)  
 12125824  
 2 (bases 1 to 10029)  
 Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,  
 Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,  
 Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,  
 Bhattacharya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,  
 Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.  
 Direct Submission  
 Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute  
 for Microbiology and Genetics, Grisebachstrasse 8, Goettingen  
 37077, Germany  
 Location/Qualifiers  
 1..10029  
 /organism="Methanosarcina mazei Goel"  
 /mol\_type="genomic DNA"  
 /strain="Goel"  
 /db\_xref="taxon:192952"  
 /complement(121..1278)  
 /gene="NM0818"  
 /complement(121..1278)  
 /gene="NM0818"  
 /EC number="2.7.2.11"  
 /codon\_start=1

**FEATURES**  
 source  
 1..10029  
 /organism="Methanosarcina mazei Goel"  
 /mol\_type="genomic DNA"  
 /strain="Goel"  
 /db\_xref="taxon:192952"  
 /complement(121..1278)  
 /gene="NM0818"  
 /complement(121..1278)  
 /gene="NM0818"  
 /EC number="2.7.2.11"  
 /codon\_start=1

**gene**  
 CDS

**misc\_feature**  
 1..10164  
 /note="BAC resource: http://bacpac.chori.org/"  
 /note="assembly\_fragment"  
 clone end:SP6  
 vector\_side:left  
 10165..10264  
 /estimated\_length=unknown  
 10265..13805  
 /note="assembly\_fragment"  
 13806..13905  
 /estimated\_length=unknown  
 13906..16621  
 /note="assembly\_fragment"  
 16622..16721  
 /estimated\_length=unknown  
 16722..60860  
 /note="assembly\_fragment"  
 60861..60780  
 /estimated\_length=unknown  
 60781..78425  
 /note="assembly\_fragment"  
 78426..78525  
 /estimated\_length=unknown  
 78526..100161  
 /note="assembly\_fragment"  
 100162..100261  
 /estimated\_length=unknown  
 100262..103084  
 /note="assembly\_fragment"  
 103085..103184  
 /estimated\_length=unknown  
 103185..105651  
 /note="assembly\_fragment"  
 105652..105751  
 /estimated\_length=unknown  
 105752..163916  
 /note="assembly\_fragment"  
 clone end:T7  
 vector\_side:right

**ORIGIN**  
 Query Match 75.2%; Score 18.8; DB 14; Length 163916;  
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAACGACAGCGAGCCCT 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 21623 AAAAAACGACAGCGAGCCCT 21644  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 10  
 CR386415/c 791 bp mRNA linear VRT 05-APR-2004  
 LOCUS Gallus gallus finished cDNA, clone CHEST231n18.  
 DEFINITION  
 ACCESSION CR386415  
 VERSION CR386415.1 GI:46239174  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Gallus gallus (chicken)  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 791)  
 Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,  
 Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,  
 Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,  
 Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,  
 Tickle,C. and Wilson,S.A.  
 Direct Submission  
 Submitted (05-APR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: chickens@bbs.ox.ac.uk  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA

**misc\_feature**  
 1..10164  
 /note="BAC resource: http://bacpac.chori.org/"  
 /note="assembly\_fragment"  
 clone end:SP6  
 vector\_side:left  
 10165..10264  
 /estimated\_length=unknown  
 10265..13805  
 /note="assembly\_fragment"  
 13806..13905  
 /estimated\_length=unknown  
 13906..16621  
 /note="assembly\_fragment"  
 16622..16721  
 /estimated\_length=unknown  
 16722..60860  
 /note="assembly\_fragment"  
 60861..60780  
 /estimated\_length=unknown  
 60781..78425  
 /note="assembly\_fragment"  
 78426..78525  
 /estimated\_length=unknown  
 78526..100161  
 /note="assembly\_fragment"  
 100162..100261  
 /estimated\_length=unknown  
 100262..103084  
 /note="assembly\_fragment"  
 103085..103184  
 /estimated\_length=unknown  
 103185..105651  
 /note="assembly\_fragment"  
 105652..105751  
 /estimated\_length=unknown  
 105752..163916  
 /note="assembly\_fragment"  
 clone end:T7  
 vector\_side:right

**ORIGIN**  
 Query Match 75.2%; Score 18.8; DB 14; Length 163916;  
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAACGACAGCGAGCCCT 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 21623 AAAAAACGACAGCGAGCCCT 21644  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 10  
 CR386415/c 791 bp mRNA linear VRT 05-APR-2004  
 LOCUS Gallus gallus finished cDNA, clone CHEST231n18.  
 DEFINITION  
 ACCESSION CR386415  
 VERSION CR386415.1 GI:46239174  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Gallus gallus (chicken)  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 791)  
 Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,  
 Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,  
 Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,  
 Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,  
 Tickle,C. and Wilson,S.A.  
 Direct Submission  
 Submitted (05-APR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: chickens@bbs.ox.ac.uk  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA

**misc\_feature**  
 1..10164  
 /note="BAC resource: http://bacpac.chori.org/"  
 /note="assembly\_fragment"  
 clone end:SP6  
 vector\_side:left  
 10165..10264  
 /estimated\_length=unknown  
 10265..13805  
 /note="assembly\_fragment"  
 13806..13905  
 /estimated\_length=unknown  
 13906..16621  
 /note="assembly\_fragment"  
 16622..16721  
 /estimated\_length=unknown  
 16722..60860  
 /note="assembly\_fragment"  
 60861

gene	/transl_table=11 /product="glutamate 5-kinase" /protein_id="AA030514.1" /db_xref="GI:20905229"			CDS	/transl_table=11 /product="Gamma-glutamyl phosphate reductase" /protein_id="AA030515.1" /db_xref="GI:20905230"														
	/translation="MTREGEQFRDVKIVIKIGTSSITRKGCDDHTRENCNIDPAFMES IAFQVSLRQGEVILVSSGAIGVGLNELGIAPKPREIPIRQAAAAGOSMLQDMS RAPSRYGMQAQILLIYEFYSRVTVLNLNRSISTLLEGVPIINENDCTCTNEIEA IFGNDKLSMAVASKIDADULLIILSDIDGLFDRNPKTHIDAKLLITVKKITPIESYG GDTPSPKVGGMRTKIKAAKICSMAGCYVIANSDVEDVILKIASGEEIGTLFLAERH IQKNRARVILARASGTVRVDAGAKAAILGKNSLLPAGVVDSVETGDFRDGVVVKLECDG KIFAKGITDYSBELIKIGVHTDQIENVLGYSNNYVIKKENIGILEBLN" complement (1416. .2759) /gene="MM0819" complement (1416. .2759) /gene="MM0819" /EC_number="1.2.1.41" /codon_start=1 /transl_table=11 /product="Gamma-glutamyl phosphate reductase" /protein_id="AA030515.1" /db_xref="GI:20905230"				/translation="MNDIESKVIKAKASIELASVSSEVKNRALEMAEALDKERI ILBANLKDLYAQLKAGKLTQALVRLKVTDSKVDGMIAGIRDVILKLPDGETLS TLELDDLLIYQSCPTGLIGVIFESRPDVVPQMSCLKSGNATIFKGSSEARESNR TIFDLVRAIESTGMEPGAFLMETREIMSLSLDAYVDLLIPRGSNEFVKFIQDN TKIPVLGHTSGIHIYDEPADPTAQVCFDAKVOYPACVNAIETLLVNRNIAEVFL PKRAEMYLKAGVRLCDGSGYSLLSEKGLSPLSRATDMSLEYNLLILSLKLVDTIK EAVDHINTFGSHHTDGIITENASRKEFTGLVDSSSVWVNASTRFADGYRGAEVG ISTNKHISRGSPVMEGLLIYKILMKGQVVADYAGKNAPKPTRKLDLKFEDVN" 3647. .4051 /gene="MM0820" 3647. .4051 /gene="MM0820" /codon_start=1 /transl_table=11 /product="conserved protein" /protein_id="AA030516.1" /db_xref="GI:20905231"				/translation="MKSFLVRADDSKVKIALHDLERYGHIOFSATPKCIEPNYADE LLVSVMGSLKSCNSAALVELNHNAGAAISRLKKIHPAHIIISPRHKMFELAGK FPKYPEDRTFNHQSPQGLDMTQEGESSNA" complement (4158. .4610) /gene="MM0821" complement (4158. .4610) /gene="MM0821" /codon_start=1 /transl_table=11 /product="conserved protein" /protein_id="AA030517.1" /db_xref="GI:20905232"				/translation="MTDUDVYRKLERIEBLLLELSAESRGAIVIVEGKRDILSMKK LGIGSPFELATRYSLFNFPSRVASLGCVELIITDWRGDLAAKUSEYFENFGVKPD LOIRNKLJASQKEIKDIESLYTYVSKRVKTSLSNSETYNLEFR" complement (4693. .5019) /gene="MM0822" complement (4693. .5019) /gene="MM0822" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA030518.1" /db_xref="GI:20905233"				/translation="WRSMSFEAAIKSLDSGIIVDIEVTPGSRSLSVSPSGNWRKRIE VKLIRNAQKGANEQLIESLAELFGICSSDIFISSGATSSKKSLILKGVSYQQAVLVP GKHLKG" 5663. .6772 /gene="MM0823" 5663. .6772 /gene="MM0823" /codon_start=1 /transl_table=11 /product="conserved protein" /protein_id="AA030519.1"		
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													
CDS						CDS													
gene						gene													



\* 3360 4096: contig of 737 bp in length  
\* 4097 gap of 100 bp  
\* 4917: contig of 721 bp in length  
\* 4918 gap of 100 bp  
\* 5017: contig of 697 bp in length  
\* 5018 gap of 100 bp  
\* 5814: contig of 743 bp in length  
\* 5815 gap of 100 bp  
\* 6557: contig of 687 bp in length  
\* 6558 gap of 100 bp  
\* 6568 7344: contig of 687 bp in length  
\* 7345 gap of 100 bp  
\* 7445 8170: contig of 726 bp in length  
\* 8171 8270: gap of 100 bp  
\* 8271 8996: contig of 726 bp in length  
\* 8997 gap of 100 bp  
\* 9087 9802: contig of 706 bp in length  
\* 9803 9902: gap of 100 bp  
\* 9903 10633: contig of 731 bp in length  
\* 10634 10733: gap of 100 bp  
\* 10734 11416: contig of 683 bp in length  
\* 11417 11516: gap of 100 bp  
\* 11517 12228: contig of 712 bp in length  
\* 12229 12328: gap of 100 bp  
\* 12329 13033: contig of 711 bp in length  
\* 13040 13139: gap of 100 bp  
\* 13140 13860: contig of 721 bp in length  
\* 13861 13960: gap of 100 bp  
\* 14674: contig of 714 bp in length  
\* 14675 14774: gap of 100 bp  
\* 14775 15489: contig of 715 bp in length  
\* 15490 15589: gap of 100 bp  
\* 15590 16320: contig of 731 bp in length  
\* 16321 16420: gap of 100 bp  
\* 16421 17106: contig of 686 bp in length  
\* 17107 17206: gap of 100 bp  
\* 17207 17903: contig of 697 bp in length  
\* 17904 18003: gap of 100 bp  
\* 18004 18740: contig of 737 bp in length  
\* 18741 18840: gap of 100 bp  
\* 18841 19547: contig of 707 bp in length  
\* 19548 19647: gap of 100 bp  
\* 19648 20308: contig of 661 bp in length  
\* 20309 20408: gap of 100 bp  
\* 20409 21124: contig of 716 bp in length  
\* 21125 21224: gap of 100 bp  
\* 21225 21939: contig of 715 bp in length  
\* 21940 22039: gap of 100 bp  
\* 22040 22699: contig of 660 bp in length  
\* 22700 22799: gap of 100 bp  
\* 22800 23552: contig of 753 bp in length  
\* 23553 23652: gap of 100 bp  
\* 23653 24371: contig of 719 bp in length  
\* 24372 24471: gap of 100 bp  
\* 24472 25173: contig of 702 bp in length  
\* 25174 25273: gap of 100 bp  
\* 25274 25974: contig of 701 bp in length  
\* 25975 26074: gap of 100 bp  
\* 26075 26789: contig of 715 bp in length  
\* 26790 26889: gap of 100 bp  
\* 26890 27611: contig of 722 bp in length  
\* 27612 27711: gap of 100 bp  
\* 27712 28427: contig of 716 bp in length  
\* 28428 28527: gap of 100 bp  
\* 28528 29251: contig of 724 bp in length  
\* 29252 29351: gap of 100 bp  
\* 29352 30053: contig of 702 bp in length  
\* 30054 30153: gap of 100 bp  
\* 30154 30849: contig of 696 bp in length  
\* 30850 30949: gap of 100 bp  
\* 30950 31536: contig of 587 bp in length  
\* 31537 31636: gap of 100 bp  
\* 31637 32369: contig of 733 bp in length  
\* 32370 32469: gap of 100 bp  
\* 32470 33188: contig of 719 bp in length

\* 33189 33288: gap of 100 bp  
\* 33289 34007: contig of 719 bp in length  
\* 34008 34107: gap of 100 bp  
\* 34108 34847: contig of 740 bp in length  
\* 34848 34947: gap of 100 bp  
\* 34948 35682: contig of 735 bp in length  
\* 35683 35782: gap of 100 bp  
\* 35783 36492: contig of 710 bp in length  
\* 36493 36592: gap of 100 bp  
\* 36593 37318: contig of 726 bp in length  
\* 37319 37418: gap of 100 bp  
\* 37419 38124: contig of 706 bp in length  
\* 38125 38224: gap of 100 bp  
\* 38225 38919: contig of 695 bp in length  
\* 38920 39019: gap of 100 bp  
\* 39020 39727: contig of 708 bp in length  
\* 39728 39827: gap of 100 bp  
\* 39828 40465: contig of 638 bp in length  
\* 40466 40565: gap of 100 bp  
\* 40566 41313: contig of 748 bp in length  
\* 41314 41413: gap of 100 bp  
\* 41414 42133: contig of 720 bp in length  
\* 42134 42233: gap of 100 bp  
\* 42234 42961: contig of 728 bp in length  
\* 42962 43061: gap of 100 bp  
\* 43062 43761: contig of 700 bp in length  
\* 43762 43861: gap of 100 bp  
\* 43862 44571: contig of 710 bp in length  
\* 44572 44671: gap of 100 bp  
\* 44673 45392: contig of 721 bp in length  
\* 45393 45492: gap of 100 bp  
\* 45493 46191: contig of 699 bp in length  
\* 46192 46291: gap of 100 bp  
\* 46292 46980: contig of 689 bp in length  
\* 46981 47080: gap of 100 bp  
\* 47081 47702: contig of 622 bp in length  
\* 47703 47802: gap of 100 bp  
\* 47803 48483: contig of 681 bp in length  
\* 48484 48583: gap of 100 bp  
\* 48584 49277: contig of 694 bp in length  
\* 49278 49377: gap of 100 bp  
\* 49378 50085: contig of 708 bp in length  
\* 50086 50185: gap of 100 bp  
\* 50186 50873: contig of 688 bp in length  
\* 50874 50973: gap of 100 bp  
\* 50974 51701: contig of 728 bp in length  
\* 51702 51801: gap of 100 bp  
\* 51802 52530: contig of 729 bp in length  
\* 52531 52630: gap of 100 bp  
\* 52631 53337: contig of 707 bp in length  
\* 53338 53437: gap of 100 bp  
\* 53438 54140: contig of 703 bp in length  
\* 54141 54240: gap of 100 bp  
\* 54241 54949: contig of 709 bp in length  
\* 54950 55049: gap of 100 bp  
\* 55050 55767: contig of 718 bp in length

Query Match 74.4%; Score 18.6; DB 14; Length 65758;  
Best Local Similarity 84.0%; Pred. No. 5.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAACACACAGCGAGCCCTCGG 25  
Db 58799 AAAAAACGATCATCGGGCCCTCGG 58823

RESULT 14  
AL845429/c AL845429 97626 bp DNA linear ROD 09-OCT-2002  
LOCUS Mouse DNA sequence from clone RP23-29004 on chromosome 2, complete  
DEFINITION sequence.  
ACCESSION AL845429  
VERSION AL845429.4 GI:23895741



KEYWORDS	HTG.	REFERENCE	1 (bases 1 to 109295)
SOURCE	Mus musculus (house mouse)	AUTHORS	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, P., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, S., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scheter, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
FEATURES	source		Unpublished
ORIGIN			2 (bases 1 to 109295)
	Query Match 74.4%; Score 18.6; DB 9; Length 97626;		Worley, K. C.
	Best Local Similarity 84.0%; Pred. No. 5.7e+02;		Direct Submission
	Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		Submitted (07-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
			3 (bases 1 to 109295)
QY	1 AAAAAACGACAGCGAGCCCTCGG 25		Worley, K. C.
Db	46018 AAAAAACGAGTCATCGGCGCCCTCGG 45994		Direct Submission
			Submitted (27-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
			Center: Baylor College of Medicine
			Center code: BCM
			Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>
			Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a>
			----- Project Information
			Center project name: GX1J
			Center clone name: CH230-418A3
			----- Summary Statistics
			Sequencing vector: Plasmid;
RESULT 15			
AC141049			
LOCUS	AC141049		
DEFINITION	Rattus norvegicus clone CH230-418A3, *** SEQUENCING IN PROGRESS		
VERSION	AC141049.1 GI:28875908		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.		

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 79468 bases at least Q40

Consensus quality: 87434 bases at least Q30

Consensus quality: 93152 bases at least Q20

Estimated insert size: 71600; sum-of-contigs estimation

Quality coverage: 1x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 54 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1064: contig of 1064 bp in length  
\* 1065 1164: gap of unknown length  
\* 1166 2209: contig of 1045 bp in length  
\* 2210 2309: gap of unknown length  
\* 2310 3382: contig of 1073 bp in length  
\* 3383 3482: gap of unknown length  
\* 3483 4805: contig of 1323 bp in length  
\* 4806 4905: gap of unknown length  
\* 4906 5908: contig of 1003 bp in length  
\* 5909 6008: gap of unknown length  
\* 6009 7048: contig of 1040 bp in length  
\* 7049 7149: gap of unknown length  
\* 7150 8241: contig of 1092 bp in length  
\* 8242 8341: gap of unknown length  
\* 8342 9357: contig of 1017 bp in length  
\* 9358 9457: gap of unknown length  
\* 9458 11038: contig of 1581 bp in length  
\* 11039 11139: gap of unknown length  
\* 11140 12277: contig of 1139 bp in length  
\* 12278 12377: gap of unknown length  
\* 12378 13987: contig of 1610 bp in length  
\* 13988 14087: gap of unknown length  
\* 14088 15318: contig of 1231 bp in length  
\* 15319 15418: gap of unknown length  
\* 15419 17008: contig of 1590 bp in length  
\* 17009 17109: gap of unknown length  
\* 17110 18313: contig of 1205 bp in length  
\* 18314 18413: gap of unknown length  
\* 18414 19671: contig of 1258 bp in length  
\* 19672 19771: gap of unknown length  
\* 19772 21223: contig of 1452 bp in length  
\* 21224 21323: gap of unknown length  
\* 21324 22637: contig of 1314 bp in length  
\* 22638 22737: gap of unknown length  
\* 22738 24082: contig of 1345 bp in length  
\* 24083 24182: gap of unknown length  
\* 24183 25905: contig of 1723 bp in length  
\* 25906 26005: gap of unknown length  
\* 26006 27289: contig of 1284 bp in length  
\* 27290 27389: gap of unknown length  
\* 27390 28814: contig of 1425 bp in length  
\* 28815 28914: gap of unknown length  
\* 28915 30635: contig of 1721 bp in length  
\* 30636 30735: gap of unknown length  
\* 30736 31813: contig of 1078 bp in length  
\* 31814 31913: gap of unknown length  
\* 31914 33042: contig of 1129 bp in length  
\* 33043 33142: gap of unknown length  
\* 33143 34227: contig of 1085 bp in length  
\* 34228 34327: gap of unknown length  
\* 34328 36307: contig of 1980 bp in length  
\* 36308 36407: gap of unknown length  
\* 36408 37796: contig of 1389 bp in length  
\* 37797 37896: gap of unknown length  
\* 37897 39456: contig of 1560 bp in length

\* 39457 39556: gap of unknown length  
\* 39557 41681: contig of 2125 bp in length  
\* 41682 41781: gap of unknown length  
\* 41782 44196: contig of 2415 bp in length  
\* 44197 44296: gap of unknown length  
\* 44297 46281: contig of 1985 bp in length  
\* 46282 46381: gap of unknown length  
\* 46382 48562: contig of 2181 bp in length  
\* 48563 48662: gap of unknown length  
\* 48663 50988: contig of 2326 bp in length  
\* 50989 51088: gap of unknown length  
\* 51089 52576: contig of 1488 bp in length  
\* 52577 52676: gap of unknown length  
\* 52677 54670: contig of 1994 bp in length  
\* 54671 54770: gap of unknown length  
\* 54771 56676: contig of 1906 bp in length  
\* 56677 56776: gap of unknown length  
\* 56777 59285: contig of 2509 bp in length  
\* 59286 59385: gap of unknown length  
\* 59386 62184: contig of 2799 bp in length  
\* 62185 62284: gap of unknown length  
\* 62285 63890: contig of 1606 bp in length  
\* 63891 63990: gap of unknown length  
\* 63991 65979: contig of 1989 bp in length  
\* 65980 66079: gap of unknown length  
\* 66080 67642: contig of 1563 bp in length  
\* 67643 67742: gap of unknown length  
\* 67743 69373: contig of 1631 bp in length  
\* 69374 69473: gap of unknown length  
\* 69474 71733: contig of 2260 bp in length  
\* 71734 71833: gap of unknown length  
\* 71834 74883: contig of 3050 bp in length  
\* 74884 74983: gap of unknown length  
\* 74984 77244: contig of 2261 bp in length  
\* 77245 77344: gap of unknown length  
\* 77345 79711: contig of 2367 bp in length  
\* 79712 79811: gap of unknown length  
\* 79812 83427: contig of 3616 bp in length  
\* 83428 83527: gap of unknown length  
\* 83528 86475: contig of 2948 bp in length  
\* 86476 86575: gap of unknown length  
\* 86576 89533: contig of 2958 bp in length  
\* 89534 89633: gap of unknown length

Query Match 74.4%; Score 18.6; DB 14; Length 109295;  
Best Local Similarity 84.0%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGACGAGCCCTCGG 25

Db 18082 AAAAAACGACGACGAGCAGG 18106

RESULT 16

AC119467 1/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC119467 Accession AC119467

Fragment Name Begin End

AC119467\_0 1 110000

AC119467\_1 100001 210000

AC119467\_2 200001 310000

AC119467\_3 300001 379464

Continuation (2 of 4) of AC119467 from base 100001 (AC119467 Rattus norvegicus clone C1

Query Match 74.4%; Score 18.6; DB 14; Length 110000;  
Best Local Similarity 84.0%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGACGAGCCCTCGG 25

Db 15657 AAAAAACGACGACGAGCAGG 15633



AP008216_141	14100001	14210000	HTG.	Oryza sativa (japonica cultivar-group)
AP008216_142	14200001	14310000	SOURCE	Oryza sativa (japonica cultivar-group)
AP008216_143	14300001	14410000	ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AP008216_144	14400001	14510000		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
AP008216_145	14500001	14610000		Ehrhartoideae; Oryzae; Oryza.
AP008216_146	14600001	14710000	REFERENCE	1 (bases 1 to 139999)
AP008216_147	14700001	14810000	AUTHORS	Buell, C.R., Yuan, Q., Moffat, K.S., Hill, J.N., Burr, P.C., Heiao, J.,
AP008216_148	14800001	14910000		Zismann, V., Pai, G., Bowman, C.L., Fujii, C.Y., VanAken, S.E.,
AP008216_149	14900001	15010000		Bowman, C.L., Craven, B., Utterback, T.R., Khalak, H., Feldblyum, T.V.,
AP008216_150	15000001	15110000		Quackenbush, J., White, O., Salzberg, S.L. and Fraser, C.M.
AP008216_151	15100001	15210000	TITLE	Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence
AP008216_152	15200001	15310000	JOURNAL	Unpublished
AP008216_153	15300001	15410000	REFERENCE	2 (bases 1 to 139999)
AP008216_154	15400001	15510000	AUTHORS	Buell, R.
AP008216_155	15500001	15610000	TITLE	Direct Submission
AP008216_156	15600001	15710000	JOURNAL	Submitted (17-DEC-1999) The Institute for Genomic Research, 9712
AP008216_157	15700001	15810000		Medical Center Dr, Rockville, MD 20850, USA
AP008216_158	15800001	15910000	REFERENCE	3 (bases 1 to 139999)
AP008216_159	15900001	16010000	AUTHORS	Buell, R.
AP008216_160	16000001	16110000	TITLE	Direct Submission
AP008216_161	16100001	16210000	JOURNAL	Submitted (05-JAN-2001) The Institute for Genomic Research, 9712
AP008216_162	16200001	16310000		Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
AP008216_163	16300001	16410000	REFERENCE	4 (bases 1 to 139999)
AP008216_164	16400001	16510000	AUTHORS	Buell, R.
AP008216_165	16500001	16610000	TITLE	Direct Submission
AP008216_166	16600001	16710000	JOURNAL	Submitted (13-MAR-2001) The Institute for Genomic Research, 9712
AP008216_167	16700001	16810000		Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
AP008216_168	16800001	16910000	REFERENCE	5 (bases 1 to 139999)
AP008216_169	16900001	17010000	AUTHORS	Buell, R.
AP008216_170	17000001	17110000	TITLE	Direct Submission
AP008216_171	17100001	17210000	JOURNAL	Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
AP008216_172	17200001	17310000		Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
AP008216_173	17300001	17410000	REFERENCE	On Jan 5, 2001 this sequence version replaced gi:12025621.
AP008216_174	17400001	17510000	AUTHORS	Address all correspondence to:rice@tigr.org
AP008216_175	17500001	17610000	TITLE	BAC clone OSJNBa0056G17 is from Oryza sativa chromosome 10
AP008216_176	17600001	17710000	JOURNAL	The orientation of the sequence is from SP6 to T7 end of the BAC
AP008216_177	17700001	17810000	COMMENT	clone.
AP008216_178	17800001	17910000		Genes were identified by a combination of several methods: Gene
AP008216_179	17900001	18010000		prediction programs including Genscan and Genscan+ (Chris Burge,
AP008216_180	18000001	18110000		http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
AP008216_181	18100001	18210000		http://genemark.biology.gatech.edu/GeneMark/), Fgenesh
AP008216_182	18200001	18310000		(http://www.softberry.com/), and GeneSplicer (Mihaela Pertea and
AP008216_183	18300001	18410000		Steven Salzberg, contact mpertea@tigr.org), searches of the
AP008216_184	18400001	18510000		complete sequence against a peptide database and the plant EST
AP008216_185	18500001	18610000		database at TIGR (http://www.tigr.org/tdb/cgi.shtml). Annotated
AP008216_186	18600001	18710000		genes are named to indicate the level of evidence for their
AP008216_187	18700001	18810000		annotation. Genes with similarity to other proteins are named after
AP008216_188	18800001	18910000		the database hits. Genes without significant peptide similarity but
AP008216_189	18900001	19010000		with EST similarity are named as unknown proteins. Genes without
AP008216_190	19000001	19110000		protein or EST similarity, that are predicted by more than two gene
AP008216_191	19100001	19210000		prediction programs over most of their length are annotated as
AP008216_192	19200001	19310000		hypothetical proteins. Genes encoding tRNAs are predicted by
AP008216_193	19300001	19410000		tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
AP008216_194	19400001	19510000		Simple repeats are identified by repeatmasker (Arian Smit,
AP008216_195	19500001	19610000		http://ftp.genome.washington.edu/RM/RepeatMasker.html).
AP008216_196	19600001	19710000		

Query Match	74.4%	Score 18.6;	DB 15;	Length 110000;	
Best Local Similarity	84.0%;	Pred. No. 5.7e+02;			
Matches 21;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;	source
QY	1	AAAAACGACACAGCGAGCCCTCGG	25		
DB	76967	AAAAAAGGACACAGCTAGCACTCAG	76943		
RESULT 18					
AC018727/c					
LOCUS	AC018727	139999 bp	DNA	linear	PLN 29-AUG-2001
DEFINITION	Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence,				
	complete sequence.				
ACCESSION	AC018727				
VERSION	AC018727.10	GI:12039362			

FEATURES	
Location/Qualifiers	
1. .139999	
/organism="Oryza sativa (japonica cultivar-group)"	
/mol_type="genomic DNA"	
/cultivar="Nipponbare"	
/sub_species="japonica"	
/db_xref="taxon:39947"	
/chromosome="10"	
/map="C405"	
/clone="OSJNBa0056G17"	
1. .213	
/genes="OSJNBa0056G17.16"	
/notes="similar to cellulose synthase 8 GB:AAF99968	
GI:9622888 (Zea mays)"	
<1. .>213	
/genes="OSJNBa0056G17.16"	

gene

mRNA

CDS	<div>&lt;1. 213</div> <div>/gene="OSJNBa0056G17.16"</div> <div>/codon_start=1</div> <div>/product="putative cellulose synthase 5-partial"</div> <div>/protein_id="AAG46166.1"</div> <div>/db_xref="GI:12039380"</div> <div>/translation="KLLGGVRFSEFWLAHLYPFKAGLMGRRGRTTIVVWVSGLVAIT</div> <div>ISLLWIAIKPPSAQNSQLGSGSEFP"</div> <div>complement(915..1010)</div> <div>/rpt_family="GGA)n"</div> <div>complement(1122..5940)</div> <div>/note="EST BE039949 from this gene"</div> <div>complement(join(1122..1267,1540..1704,1788..1867,</div> <div>2028..2131,2343..2544,3960..4113,4204..&gt;5940))</div> <div>/gene="OSJNBa0056G17.1"</div> <div>complement(join(4045..4113,4204..5940))</div> <div>/codon_start=1</div> <div>/product="unknown protein"</div> <div>/protein_id="AAG46167.1"</div> <div>/db_xref="GI:12039381"</div> <div>/translation="MLHATRTPTRLSAAAAAFTTTPRSHPPPPRSLPRILVDA</div> <div>TVSRCPDALAITFFLMWCARCPAYFHPSPDFRLLPAAARLASRLRTAPAILHQHRL</div> <div>GCPIRPTLLRLLYWRCGIYPLVQLFQDOMLGFHFNAPARNVLDVLLRTHHH</div> <div>SALCFLRDNPENLYTAYLLHLCRSGNWPGVACFLAMHLHQGLPSAASINLPAC</div> <div>CTKNAISLQLLGLFTLVSGYQLTSAMTICLARKCRGKJDEAIRMLAKMLASGSP</div> <div>PTVYTPTRALYRAGRHDIATELFAFSGNSTNCSPDLVLHNVLDMDTKDKRYDAAL</div> <div>GVLMLHESQMKPDATYTLSTLRALHLSRVNSLLPRLFLDSADIYDLVACNSVLNAL</div> <div>RKSGPQAVQFIFINIKCDIRPDGVSYGVLGDSLCOLGRIDHAINVHISIVSSPDS</div> <div>NAYVHAAILCGIVKKGHNEMALINEAVRONVALDVCYTVVLHGLQAHILIEACM</div> <div>LFDRKSGMASTCTYNIIMRLGRCTRTHALKWFLREMECSDEVMSISYNILVVF</div> <div>LKLQHISATDALANIELHSDSDTNDLLACAS"</div> <div>5604..5785</div> <div>/rpt_family=""(CGGGGG)"</div> <div>5835..5921</div> <div>/rpt_family=""(CGG)n"</div> <div>complement(8345..9535)</div> <div>/note="OSJNBa0056G17.17"</div> <div>/notes="predicted by genemark hmm"</div> <div>complement(&lt;8345..&gt;9535)</div> <div>/gene="OSJNBa0056G17.17"</div> <div>complement(8345..9535)</div> <div>/gene="OSJNBa0056G17.17"</div> <div>/codon_start=1</div> <div>/product="hypothetical protein"</div> <div>/protein_id="AAG46168.1"</div> <div>/db_xref="GI:12039382"</div> <div>/translation="MADSHLDGVPMARKTTVSSSSSLNTAFLLSMVATNLLSLYHLST</div> <div>RASAPPLLLPAHQAOQDEQLLQTLTAIRATVSQLNHLRSSTPPPPPPPELLIYS</div> <div>RLAPLASASHPDLLHREMSYTFPSPCDDALSLAPULLRGCHPLPERRCFSSSI</div> <div>SSSHNLNLTDFPSLPDSAVRWPGAKTFSCLPSPGLDLARTEARFLRAEGP</div> <div>LDLTVPQLRLASLSRAGTIRGLDVGCGTGTIAARLKRAANATVLTMTTINLGAYPE</div> <div>AAAARGVPLHAPLQORFPVGDATDMLVRAGHAVNRWIPEAALEFLWYDMDRLVRPRG</div> <div>LLWVDHFWCRRPDLAAVYQPMRLRRLGYKTLKWAVADKTTPTTAPGAKHHDVYLTAL</div> <div>LQKP"</div> <div>9206..9290</div> <div>/rpt_family=""(CGGGGG)"</div> <div>complement(9691..13329)</div> <div>/gene="OSJNBa0056G17.2"</div> <div>/note="EST D47052, AU096093, AU075883 from this gene"</div> <div>complement(join(&lt;9691..10099,10826..10955,11285..11459,</div> <div>11659..11768,11903..12018,12376..12637,12766..13075,</div> <div>13214..&gt;13329))</div> <div>/gene="OSJNBa0056G17.2"</div> <div>complement(join(9913..10099,10826..10955,11282..11459,</div> <div>11659..11768,11903..12018,12376..12637,12766..13075,</div> <div>13214..13297))</div> <div>/gene="OSJNBa0056G17.2"</div> <div>/codon_start=1</div> <div>/product="unknown protein"</div> <div>/protein_id="AAG46171.1"</div> <div>/db_xref="GI:12039385"</div>	<div>repeat_region</div> <div>gene</div> <div>mRNA</div> <div>CDS</div>
repeat_region	<div>complement(10278..10350)</div> <div>/rpt_family="AT_rich"</div> <div>complement(12207..12227)</div> <div>/rpt_family="AT_rich"</div> <div>complement(14187..16368)</div> <div>/gene="OSJNBa0056G17.15"</div> <div>/note="highly similar to cholinephosphate</div> <div>cytidyltransferase GB:CAAY0317 GI:1657382 (Pisum</div> <div>sativum); EST C26593 from this gene"</div> <div>complement(join(&lt;14187..14282,14405..14515,14635..14703,</div> <div>14841..14996,15085..15127,15235..15330,15805..15918,</div> <div>16059..&gt;16368))</div> <div>/gene="OSJNBa0056G17.15"</div> <div>complement(join(14187..14282,14405..14515,14635..14703,</div> <div>14841..14996,15085..15127,15235..15330,15805..15918,</div> <div>16059..16243))</div> <div>/gene="OSJNBa0056G17.15"</div> <div>/codon_start=1</div> <div>/product="putative cholinephosphate cytidyltransferase"</div> <div>/protein_id="AAG46173.1"</div> <div>/db_xref="GI:12039387"</div> <div>/translation="MARVSHRKRTTRAPSNSKPTSQTPPPPELTDPRVVRVADGI</div> <div>FDLHFHARALEQALFPNTYLLVGCNDELNTRYKGTVMOTDERYESLRHCKWV</div> <div>DEVIPDAVWLTOEFLDKHODYVAHDALPYADTSGAANDVYEFVKIKGKETKRTD</div> <div>GVSTDLIMRLIKDYNQYMRNLARGYTRKDLGYSYVKEKQLQVNMKINKLRETVKAH</div> <div>QEKLTVAKTAGINHEWLANADRWVAGFLKFEHCHNMTAIGRIQEKLRQTSR</div> <div>GIIGLMQPPVAA"</div> <div>complement(16596..16625)</div> <div>/rpt_family=""(TAA)n"</div> <div>complement(16673..16698)</div> <div>/rpt_family=""(CATCG)n"</div> <div>complement(17509..20874)</div> <div>/gene="OSJNBa0056G17.30"</div> <div>/note="similar to 4 coumarate CoA ligase GB:AAC24503</div> <div>GI:3258635 (Populus tremuloides)"</div> <div>complement(join(&lt;17509..17640,18143..18245,18370..18437,</div> <div>18565..18722,19375..19564,19831..&gt;20874))</div> <div>/gene="OSJNBa0056G17.30"</div> <div>complement(join(17509..17640,18143..18245,18370..18437,</div> <div>18565..18722,19375..19564,19831..20874))</div>	<div>repeat_region</div> <div>repeat_region</div> <div>gene</div> <div>mRNA</div> <div>CDS</div>
repeat_region	<div>Query Match 74.4%; Score 18.6; DB 15; Length 139999;</div> <div>Best Local Similarity 84.0%; Pred. No. 5.6e+02; Mismatches 4; Indels 0; Gaps 0;</div> <div>Matches 21; Conservative 0;</div> <div>QY 1 AAAAAACACACACAGCGCCCTCGG 25</div> <div>     </div> <div>Db 31636 AAAAAAGACACAGCTAGACTCAG 31612</div>	<div>repeat_region</div> <div>repeat_region</div> <div>gene</div> <div>mRNA</div> <div>CDS</div>
repeat_region	<div>RESULT 19</div> <div>AC158566/c</div> <div>LOCUS</div> <div>AC158566</div> <div>DEFINITION Mus musculus chromosome 8, clone RP24-149112, complete sequence.</div> <div>AC158566</div> <div>ACCESSION</div> <div>AC158566.8</div> <div>VERSION</div> <div>GI:71361812</div> <div>KEYWORDS</div> <div>HTG.</div> <div>SOURCE</div> <div>Mus musculus (house mouse)</div> <div>ORGANISM</div> <div>Mus musculus</div> <div>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</div> <div>Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;</div> <div>Sciurognathi; Muroidea; Muridae; Murinae; Mus.</div> <div>REFERENCE</div> <div>1 (bases 1 to 156342)</div> <div>AUTHORS</div> <div>Birren,B., Nusbaum,C. and Lander,E.</div>	<div>repeat_region</div> <div>gene</div> <div>mRNA</div> <div>CDS</div>

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mus musculus chromosome 8, clone RP24-149112  
Unpublished  
2 (bases 1 to 156342)  
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,  
Choepl, Y., Collumore, A., Cook, A., Cooke, P., Corum, B.,  
DeArellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,  
Galgan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-MAR-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 156342)

REFERENCE  
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,  
Choepl, Y., Collumore, A., Cook, A., Cooke, P., Corum, B.,  
DeArellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,  
Galgan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (02-JUN-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 156342)  
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J.,  
Choepl, Y., Collumore, A., Cook, A., Cooke, P., Corum, B.,  
DeArellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,  
Galgan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (28-JUL-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA  
On Jul 28, 2005 this sequence version replaced gi:66865067.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## COMMENT

----- Genome Center  
Center: Broad Institute of MIT and Harvard  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: L31955  
Center clone name: 149\_I\_12

FEATURES  
source

----- Location/Qualifiers  
1..156342  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="8"  
/map="8"  
/clone="RP24-149112"  
3222..2360 /rpt\_family="AT\_rich"  
4476..4499 /rpt\_family="AT\_rich"  
5021..5061 /rpt\_family="AT\_rich"  
complement(5158..6310) /rpt\_family="TATATG)n"  
6341..6404 /rpt\_family="RMR12"  
6405..6687 /rpt\_family="MTB"  
7121..8395 /rpt\_family="Lx8"  
8390..14584 /rpt\_family="L1VL2"  
14592..15204 /rpt\_family="L1\_MM"  
15191..16139 /rpt\_family="L1"  
16142..16184 /rpt\_family="L1"  
16740..16927 /rpt\_family="AT\_rich"  
18503..18546 /rpt\_family="B2\_Mm2"  
19421..19575 /rpt\_family="(CA)n"  
19605..21143 /rpt\_family="B3A"  
21265..21515 /rpt\_family="L1\_MM"  
22058..22282 /rpt\_family="L1ME"  
22724..22800 /rpt\_family="B3"  
complement(23142..23259) /rpt\_family="AT\_rich"  
complement(23346..23635) /rpt\_family="ID\_B1"  
24022..24064 /rpt\_family="Lx2"  
24238..24315 /rpt\_family="AT\_rich"  
24740..24783 /rpt\_family="MYSERV"  
24905..24957 /rpt\_family="(TCC)n"

```
repeat_region /rpt family=" (TC)n"
24958 .25002
/rpt family=" (CA)n"
25058 .25502
/rpt family="RNER20"
26109 .26257
/rpt family="RSINE1"
26280 .26330
/rpt family=" (CA)n"
complement(27490. .27826)
/rpt family="IAPLTR1_MM"
complement(27827. .32869)
/rpt family="IAPLTR1_MM-int"
complement(32870. .33012)
/rpt family="MUT1E1a"
complement(33013. .34208)
/rpt family="IAPLTR1_MM-int"
complement(34209. .34545)
/rpt family="IAPLTR1_MM"
34745 .34928
/rpt family="L1"
35655 .35683
/rpt family=" (GAAA)n"
35686 .35741
/rpt family=" (GGGA)n"
36660 .36822
/rpt family=" (TCC)n"
37783 .38108
/rpt family="ORR1A2"
complement(38140. .38578)
/rpt family="L1_MM"
38903 .39103
/rpt family="B3"
39291 .39319
/rpt family="AT_rich"
39505 .39543
/rpt family=" (TCTCTG)n"
complement(39590. .39665)
/rpt family="B4a"
39820 .40214
/rpt family="MTA"
40410 .40486
repeat_region
```

Query Match 74.4%; Score 18.6; DB 9; Length 156342;  
Best Local Similarity 84.0%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 32414 AAAAAACGAATCATCGGCGCCTCGG 32390

```
RESULT 20
AC160549 AC160549 157085 bp DNA linear HTG 10-JUN-2005
LOCUS Mus musculus chromosome 3 clone RP23-276K18 map 3, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC160549
VERSION AC160549.4 GI:67460011
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 157085)
REFERENCE Birren,B., Nusbaum,C. and Lander,E.
AUTHORS Mus musculus chromosome 3, clone RP23-276K18
TITLE Unpublished
JOURNAL
REFERENCE Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
2 (bases 1 to 157085)
AUTHORS Anderson,M., Anderson,S., Atachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe
```

```
l,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 157085)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Atachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe
```

l,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.

Center: Broad Institute of MIT and Harvard  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: L32177  
Center clone name: 276\_K\_18

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 115321 115420: contig of 115320 bp in length  
\* 115321 115420: gap of unknown length  
\* 115421 157085: contig of 41665 bp in length.  
Location/Qualifiers  
1. .157085

FEATURES  
source



```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-86013"
/clone_lib="RPCI-23"
/map="3"
/clone="RP23-276K18"
/clone_lib="RPCI-23 Female Mouse BAC"
115321..115420
/estimated_length=unknown

gap
Query Match 74.4%; Score 18.6; DB 14; Length 157085;
Best Local Similarity 84.0%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

ORIGIN
QY 1 AAAAAACGACACGAGCCCTCGG 25
Db 72877 AAAAAACGCGCCATCGAGCCCTCGG 72901
|||||
1 AAAAAACGACACGAGCCCTCGG 25

RESULT 21
AL805943 192561 bp DNA linear ROD 17-DEC-2004
LOCUS Mouse DNA sequence from clone RP23-86013 on chromosome 4 Contains a
DEFINITION opsin 1 (cone pigments) (Opn1) pseudogene, complete sequence.
ACCESSION AL805943
VERSION AL805943.6 GI:46406414
KEYWORDS HTG; Opn1; opsin.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192561)
North,P., Leaves,N., Greystrom,J., Coppola,M., Manjunath,S.,
Russell,E., Smith,M., Strachan,G., Tofts,C., Boal,E., Cobley,V.,
Hunter,G., Kimberley,C., Thomas,D., Cave-Berry,L., Weston,P. and
Botcherby,M.R.M.
Direct Submission
Submitted (16-DEC-2004) Mouse Sequencing Group, HGMP-RC, Hinxton,
Cambridge, CB10 1SP, UK. E-mail enquiries:- mrbotche@hgmp.mrc.ac.uk
or phn0rth@hgmp.mrc.ac.uk
HGMP-RC part of the UK Mouse Sequencing Consortium
On Apr 19, 2004 this sequence version replaced gi:28172099.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseq@har.mrc.ac.uk
-----
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-86013 is
from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
Location/Qualifiers
1..192561
/organism="Mus musculus"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-86013"
/clone_lib="RPCI-23"
175628..176030
/locus_tag="RP23-86013.1-001"
/pseudo
175628..176030
/locus_tag="RP23-86013.1-001"
/notes="match: Proteins: SW:Q35599 Sw:Q35599 Tr:Q98UH7 Tr:Q9JLS7 Tr:Q9QY62
Tr:Q864C4 Tr:Q864C9 Tr:Q98UH7 Tr:Q9JLS7 Tr:Q9QY62
Tr:Q9Y152"
/pseudo
/codon_start=1
/product="opsin 1 (cone pigments) (Opn1) pseudogene"

ORIGIN
Query Match 74.4%; Score 18.6; DB 9; Length 192561;
Best Local Similarity 84.0%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25
Db 191609 AAAAAATGACCCAGCTAGCCCTCTG 191633
|||||
1 AAAAAACGACACGAGCCCTCGG 25

RESULT 22
AC122175 196097 bp DNA linear PRI 18-SEP-2003
LOCUS Pan troglodytes clone rp43-99g12, complete sequence.
DEFINITION AC122175
ACCESSION AC122175
VERSION AC122175.9 GI:34849974
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Euarchothelires; Primates; Catarrhini;
Hominidae; Pan.
1 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Pan troglodytes BAC Clone rp43-99g12
Unpublished
2 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (23-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (02-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (29-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (22-OCT-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 196097)
Shi,R., Fu,Y., Zhou,L., Wu,J., Shaul,S., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (23-OCT-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

8 (bases 1 to 196097)
Shi R., Fu Y., Zhou L., Wu J., Shaul S., Eichler E. and Roe B.A.
Direct Submission
Submitted (23-JAN-2003) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

8 (bases 1 to 196097)
Shi R., Fu Y., Zhou L., Wu J., Shaul S., Eichler E. and Roe B.A.
Direct Submission
Submitted (18-SEP-2003) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT
On Sep 18, 2003 this sequence version replaced gi:27877221.

----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

-----
Location/Qualifiers
1. 196097
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="rp43-99g12"
/clone_lib="RPCI - 43 Male Chimpanzee BAC Library"

FEATURES
source
Query Match 74.4%; Score 18.6; DB 8; Length 196097;
Best Local Similarity 84.0%; Pred. NO. 5.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

ORIGIN
1 AAAAAAGCAGACAGCGAGCCCTCGG 25
||||| ||||| ||||| |||||
Db 152949 AAAAAAGCAGACAGCGAGCCCTCGG 152973

RESULT 23
AC162627
LOCUS
DEFINITION
Bos taurus clone CH240-124D18, *** SEQUENCING IN PROGRESS ***, 18
unordered pieces.
AC162627
AC162627.2 GI:68301357
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Bos taurus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 214864)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

```

```

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuheva, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shivatsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneid, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosai, U.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 214864)
Worley, K.C.
Direct Submission
Submitted (31-MAY-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 214864)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:66796257.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FHNA
Center clone name: CH240-124D18
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 189868 bases at least Q40
Consensus quality: 191139 bases at least Q30
Consensus quality: 192213 bases at least Q20
Estimated insert size: 194905; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces

```





Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (29-Oct-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 29, 2003 this sequence version replaced gi:37515300.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WITBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L22279  
 Center clone name: 9\_p\_8

#### FEATURES

##### Location/Qualifiers

```

1. .218746
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="3"
  /map="3"
  /clone="RP23-9P8"
  /clone_lib="RPC1-23 Female Mouse BAC"
  1. .459
  /rpt_family="MWERVL_LTR"
  1. .143
  /note="wgs end extension"
  clone_end:SP6"
  /note="single clone coverage"
  complement(39. .107)
  complement(68. .80)
  /note="<30 qual single clone coverage"
  144. .149
  /note="clone boundary"
  clone_end:SP6
  site:ECORI"
  complement(732. .6710)
  /rpt_family="L1F"
  complement(4845)
  /note="some subclones have A, not G"
  complement(5614. .5642)
  /note="single clone coverage"
  complement(6711. .6949)
  /rpt_family="L1F"
  6957. .7116
  /rpt_family="Tigger7"
  8101. .8205
  /rpt_family="MTD"
  8572. .8700
  /rpt_family="Lx8"
  complement(10095. .10273)
  /rpt_family="B2_Mm2"
  10306. .10348
  /rpt_family="T-rich"
  complement(10354. .10476)
  /rpt_family="B1_MM"
  10627. .10657
  /rpt_family="(T)n"
  10943. .10998
  /rpt_family="CT-rich"
  complement(10999. .11116)
  /rpt_family="B1_MM"
  complement(11145. .11216)
  /rpt_family="ID_B1"
  complement(11498. .11622)
  /rpt_family="B1F"
  14610. .14793

```

```

repeat_region      /rpt_family="B2_Mm2"
                    complement(15055. .15243)
repeat_region      /rpt_family="B3A"
                    15308. .15340
repeat_region      /rpt_family="AT_rich"
                    complement(15720. .15856)
repeat_region      /rpt_family="B1_MM"
                    complement(15728. .15922)
repeat_region      /rpt_family="ID_B1"
                    16683. .16948
repeat_region      /rpt_family="B4A"
                    17330. .17731
repeat_region      /rpt_family="RLTR9B"
                    17785. .17935
repeat_region      /rpt_family="RSINE1"
                    17982. .18192
repeat_region      /rpt_family="B3"
                    18456. .18753
repeat_region      /rpt_family="LIM3"
                    19692. .19898
repeat_region      /rpt_family="B3"
                    20258. .20336
repeat_region      /rpt_family="B1F"
                    20337. .20358
repeat_region      /rpt_family="ID_B1"
                    21130. .21160
repeat_region      /rpt_family="AT_rich"
                    21668. .21715
repeat_region      /rpt_family="(TCTA)n"
                    21716. .21738
repeat_region      /rpt_family="(CTTA)n"
                    complement(21789. .21901)
repeat_region      /rpt_family="B4"
                    23083. .23111
repeat_region      /rpt_family="(CA)n"
                    23138. .23377
repeat_region      /rpt_family="LIM4"
                    24421. .25333
repeat_region      /rpt_family="LIM4"
                    25328. .31558
repeat_region      /rpt_family="L1_MM"
                    31631. .31973
repeat_region      /rpt_family="MTC"
                    32046. .32635
repeat_region      /rpt_family="LIM4"
                    32636. .32832

```

Query Match 74.4%; Score 18.6; DB 9; Length 218746;  
 Best Local Similarity 84.0%; Pred. No. 5.Se+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25  
 Db 176526 AAAAAACGCGCCATCGAGCCCTCGG 176550

#### RESULT 26

AC138614  
 LOCUS 230451 bp DNA linear HTG 25-FEB-2003  
 DEFINITION Mus musculus chromosome UNK clone RP23-421A7, WORKING DRAFT  
 SEQUENCE, 5 unordered pieces.  
 AC138614  
 ACCESSION AC138614.2 GI:27777615  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 REFERENCE 1 (Bases 1 to 230451)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished

```
REFERENCE
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (12-JAN-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (25-FEB-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
COMMENT      On Jan 17, 2003 this sequence version replaced gi:27657600.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BA0421A07
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 228476 bases at least Q40
Consensus quality: 228826 bases at least Q30
Consensus quality: 229089 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 235563; sum-of-contigs
Quality coverage: 11.98 in Q20 bases; agarose-fp
Quality coverage: 11.18 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1343: contig of 1343 bp in length
* 1344: gap of unknown length
* 1444: contig of 1512 bp in length
* 2956: gap of unknown length
* 3056: contig of 23876 bp in length
* 26931: gap of unknown length
* 27032: contig of 60137 bp in length
* 87168: gap of unknown length
* 87269: contig of 143183 bp in length.

FEATURES
source
  1..230451
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="UNK"
    /clone="RP23-421A7"
  1..1343
    /notes="assembly_name:Contig6"
  1344..1443
    /estimated_length=unknown
  1444..2955
    /notes="assembly_name:Contig21"
  2956..3055
    /estimated_length=unknown
  3056..26931
    /note="assembly_name:Contig22"
  26932..27031
    /estimated_length=unknown
  27032..87168
    /notes="assembly_name:Contig23"
  87169..87268

-----
Query Match      74.4%; Score 18.6; DB 14; Length 230451;
Best Local Similarity 84.0%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
Db 202023 AAAAAACGATCATCGGCGCCCTCGG 202047

-----
RESULT 27
AC119312/c
LOCUS      Rattus norvegicus clone CH230-342E8, WORKING DRAFT SEQUENCE.
DEFINITION AC119312
ACCESSION  AC119312.5 GI:24635493
VERSION    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridea; Muridae; Rattus.
REFERENCE  1 (bases 1 to 236746)
AUTHORS    Muzny,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,D., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
```







* 64451	64550: gap of 100 bp		
* 64551	67608: contig of 3058 bp in length	misc_feature	/clone_lib="RPCI-23"
* 67609	67708: gap of 100 bp		1. .15346
* 67709	73422: contig of 6234 bp in length	misc_feature	/note="assembly fragment:00842"
* 73443	74042: gap of 100 bp		fragment_chain:1"
* 74043	91966: contig of 17954 bp in length	misc_feature	15447. .43658
* 91997	92096: gap of 100 bp		/note="assembly fragment:01132"
* 92097	94347: contig of 2251 bp in length	misc_feature	fragment_chain:1"
* 94348	94447: gap of 100 bp		43759. .64450
* 94448	104001: contig of 9554 bp in length	misc_feature	/note="assembly fragment:00962"
* 104002	104101: gap of 100 bp		fragment_chain:1"
* 104102	109235: contig of 5134 bp in length	misc_feature	64551. .67608
* 109236	109335: gap of 100 bp		/note="assembly fragment:00299"
* 109336	114190: contig of 4855 bp in length	misc_feature	fragment_chain:1"
* 114191	114290: gap of 100 bp		67709. .73942
* 114291	124963: contig of 10873 bp in length	misc_feature	/note="assembly fragment:00478"
* 124964	125063: gap of 100 bp		fragment_chain:1"
* 125064	129213: contig of 4150 bp in length	misc_feature	74043. .91996
* 129214	129313: gap of 100 bp		/note="assembly fragment:00728"
* 129314	132069: contig of 2756 bp in length	misc_feature	fragment_chain:1"
* 132070	132169: gap of 100 bp		92097. .94347
* 132170	136359: contig of 4190 bp in length	misc_feature	/note="assembly fragment:00061"
* 136360	136459: gap of 100 bp		fragment_chain:2"
* 136460	144197: contig of 7738 bp in length	misc_feature	94448. .104001
* 144198	144297: gap of 100 bp		/note="assembly fragment:00377"
* 144298	146301: contig of 2004 bp in length	misc_feature	fragment_chain:2"
* 146302	146401: gap of 100 bp		104102. .109235
* 146403	158833: contig of 12432 bp in length	misc_feature	/note="assembly fragment:00151"
* 158834	159333: gap of 100 bp		fragment_chain:2"
* 159334	166350: contig of 7417 bp in length	misc_feature	109336. .114190
* 166351	166450: gap of 100 bp		/note="assembly fragment:00138"
* 166451	168700: contig of 2250 bp in length	misc_feature	fragment_chain:2"
* 168701	168800: gap of 100 bp		114291. .124963
* 168801	171153: contig of 2353 bp in length	misc_feature	/note="assembly fragment:00524"
* 171154	171253: gap of 100 bp		fragment_chain:2"
* 171254	174131: contig of 2878 bp in length	misc_feature	125064. .129213
* 174132	174231: gap of 100 bp		/note="assembly fragment:00257"
* 174232	176746: contig of 2515 bp in length	misc_feature	fragment_chain:3"
* 176747	176846: gap of 100 bp		129314. .132069
* 176847	195914: contig of 19068 bp in length	misc_feature	/note="assembly fragment:00077"
* 195915	196014: gap of 100 bp		fragment_chain:3"
* 196015	198745: contig of 2731 bp in length	misc_feature	132170. .136359
* 198746	198845: gap of 100 bp		/note="assembly fragment:00192"
* 198846	200865: contig of 2020 bp in length	misc_feature	fragment_chain:3"
* 200866	200965: gap of 100 bp		136460. .144197
* 200966	203893: contig of 2928 bp in length	misc_feature	/note="assembly fragment:00276"
* 203894	208244: contig of 4251 bp in length		fragment_chain:3"
* 208245	208344: gap of 100 bp	misc_feature	144298. .146301
* 208346	213055: contig of 4711 bp in length		/note="assembly fragment:00069"
* 213056	213155: gap of 100 bp	misc_feature	fragment_chain:4"
* 213156	218757: contig of 5602 bp in length		146402. .158833
* 218758	218857: gap of 100 bp	misc_feature	/note="assembly fragment:00573"
* 218858	222834: contig of 3977 bp in length		fragment_chain:4"
* 222835	222934: gap of 100 bp	misc_feature	158934. .166350
* 222935	228500: contig of 5566 bp in length		/note="assembly fragment:00322"
* 228501	228600: gap of 100 bp	misc_feature	fragment_chain:4"
* 228601	233666: contig of 5066 bp in length		166451. .168700
* 233667	233766: gap of 100 bp	misc_feature	/note="assembly fragment:00043"
* 233767	242489: contig of 8723 bp in length		fragment_chain:5"
* 242490	242589: gap of 100 bp	misc_feature	168801. .171153
* 242590	244863: contig of 2274 bp in length		/note="assembly fragment:00115"
* 244864	244963: gap of 100 bp	misc_feature	fragment_chain:5"
* 244964	248270: contig of 3307 bp in length		171254. .174131
* 248271	248370: gap of 100 bp	misc_feature	/note="assembly fragment:00086"
* 248371	258567: contig of 10197 bp in length.		fragment_chain:6"
	Location/Qualifiers		174232. .176746
	1. .258567		/note="assembly fragment:00105"
	/organism="Mus musculus"		fragment_chain:6"
	/mol_type="Genomic DNA"		176847. .195914
	/db_xref="taxon:10090"		/note="assembly fragment:00636"
	/chromosome="4"		fragment_chain:7"
	/clone="RP23-30108"		196015. .198745
			/note="assembly fragment:00033"

FEATURES  
source

```

misc_feature 198846..200865
/note="assembly_fragment:00056"

Query Match 74.4%; Score 18.6; DB 14; Length 258567;
Best Local Similarity 84.0%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25
||||| ||||| ||||| ||||| |||||
Db 56557 AAAAAATGACCGAGTAGCCCTCTG 56581

RESULT 29
AE017122/c
LOCUS AE017122
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 10, section 76 of
77 of the complete sequence.
ACCESSION AE017122
VERSION AE017122.1
KEYWORDS GI:31433648
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 300029)

REFERENCE
1. The Rice Chromosome 10 Sequencing Consortium
In-depth view of structure, activity, and evolution of rice
chromosome 10
Science 300, 1566-1569 (2003)
2 (bases 1 to 300029)
Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
Direct Submission
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant
peptide similarity but with EST similarity are named as unknown
proteins. Genes without protein or EST similarity, that are
predicted by more than two gene prediction programs over most of
their length are annotated as hypothetical proteins. Genes
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
http://genome.wustl.edu/eddy/tRNAscan-SE/).

FEATURES
source
1..300029
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"
<1..64740
/note="Chromosome Sequence Derivation: nucleotide sequence
in this region was derived from BAC clone OSJNBa0003019
(GB:AC060755)."
<7418..>8247
/locus_tag="OSJNBa0003019.23"
/note="similar to RING H2 zinc finger protein ATLS
GB:AAD33583 GI:4928401 (Arabidopsis thaliana)"
join(<7418..7511,7637..>8247)
/locus_tag="OSJNBa0003019.23"
join(7418..7511,7637..8247)
/locus_tag="OSJNBa0003019.23"
/codon_start=1
/product="putative zinc finger protein"
/protein_id="AAP55133.1"
/db_xref="GI:31433649"

gene
/locus_tag="OSJNBa0003019.20"
/note="similar to transcription factor MYC7B GB:AAD15818
GI:4321762 (Zea mays)"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
/codon_start=1
/product="putative MYC transcription factor"

translation="MSCRRGVAFPALATACMPPIIRKPHGGHHTTTPRHAPMAAPTRR
LLLLNHEPTTTTSSSSSSSSQSFPTLLPVFLVLLCLSLFPLVRLDLHLHLSLCURR
RRLRLRHGHDHSDSSVSQSSQSFINAQRPKPPGLDPAIILASFTPLRFKASAAPEFC
AVCLSDFAAGDALRLTLVCRHAFHTPCIDSWLRAHTTCPCVCRSDLDADAPAPRHEDP
AAVAVDVRCDREGGAG"
complement(<11042..>12329)
/locus_tag="OSJNBa0003019.19"
/note="similar to replication factor C (40 kDa subunit)
GB:AAC71957 GI:3845290 (Plasmodium falciparum)"
complement(join(<11042..11089,11194..11250,11370..11402,
11498..11608,11718..>12329))
/locus_tag="OSJNBa0003019.19"
complement(join(11042..11089,11194..11250,11370..11402,
11498..11608,11718..12329))
/locus_tag="OSJNBa0003019.19"
/codon_start=1
/product="putative replication factor"
/protein_id="AAP55134.1"
/db_xref="GI:31433650"
/translation="MAPPATTLSDALDHRDRDARRYRTFLPPSQVLGLHRLGFAFACI
PRSGAPLTRPTAPSPVHPQPPREAPQVANAARERPLREEREAAAAAANK
RTHDHTPAESKAEAAESEPOVSAGDHSTOSSDHSASAVAVRGGSIWVRVLPKSAVMV
DSPQTSSESTPTAEDKYIWADKYRNFLNDFICNDAALELYNQVTAQECNHIIIFEPG
TSVGRKSWVALIWDAPATDNLIKIEQTKRFELKGEIAKHIDIRVKLISSHHVERYMR
LFLVLRMMKQ"
complement(<19031..>20779)
/locus_tag="OSJNBa0003019.3"
/note="EST D40782, AU070707, AU162030, AT003534 from this
gene"
complement(join(<19031..19917,20011..20253,20720..>20779))
/locus_tag="OSJNBa0003019.3"
complement(join(19456..19917,20011..20253,20720..>20779))
/locus_tag="OSJNBa0003019.3"
/codon_start=1
/product="unknown protein"
/protein_id="AAP55135.1"
/db_xref="GI:31433651"
/translation="MAIDHESPFKELRNKRRIMGGGPEPEEEVAHGEQMPRWLS
PLLSASFQCKVHADSHRSGECNMFCLCAADADAAAAALCLAHNRDHTTQI
RSSVHDVIRSDIORFMDIGVQTVINSRVVFLNRPQKAGKAGAVANICEVCSR
KLSFTPTPPPTLPPKRGKIPHRAPFGLIVEY"
complement(<26603..>30151)
/locus_tag="OSJNBa0003019.4"
/note="EST C19549 from this gene"
complement(join(<26603..26710,26783..26884,27179..27301,
28054..28314,28504..28933,29724..29842,30071..>30151))
/locus_tag="OSJNBa0003019.4"
complement(join(26603..26710,26783..26884,27179..27301,
28054..28314,28504..28933,29724..29842,30071..30151))
/locus_tag="OSJNBa0003019.4"
/codon_start=1
/product="unknown protein"
/protein_id="AAP55136.1"
/db_xref="GI:31433652"
/translation="MATDSSSSAADKNTVFRKLRAKSDNMKFCDCNAKNTWASVTY
GVFLICDCSAVHRSLGVHVFVRSTNLDWTPEQLKMYVGNNAQAQFQKHGWTDG
KTEAKYTSRADLYRQLAKQVAKNSTEDGNNSSWPSSPVAAQPTNOADAIPLDLKLA
EASKVANKEPEVTRSPRAPHSPKPIVAKPGKGTGLGARKLTSKPNSLYEQ
KEPELAPLPVPTENSTAKSHTSRFEVETVTSAGSSEENQVIGHVAPPKSSNF
KEFGMDSGYHKKSAPGSKVQVEESSEARQFSNAKSISSSQFQDOQSFPEKAQVSL
QKFGSGSAISSADLFQHTPTNSSNVDSASDLNRLSFQASQDLSSIKNMAGETGKLT
SLASNTMSDLDRL"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
/note="similar to transcription factor MYC7B GB:AAD15818
GI:4321762 (Zea mays)"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
<46888..>48954
/locus_tag="OSJNBa0003019.20"
/codon_start=1
/product="putative MYC transcription factor"

```

QY	1	AAAAAACGACACAGCGAGCCTCGG 25	
Db	286923	AAAAAAGGACACAGCTAGCACTCAG 286899	
RESULT 30			
BX322621		142359 bp	DNA linear HTG 09-JUL-2004
LOCUS			
DEFINITION	Danio rerio clone RP71-80B24, WORKING DRAFT SEQUENCE, 4 unordered pieces.		
ACCESSION	BX322621		
VERSION	BX322621.19	GI:50199216	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
REFERENCE	1	(bases 1 to 142359)	
AUTHORS	Lloyd,D.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JUL-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 10, 2004 this sequence version replaced gi:49658399.		
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: zfish-help@sanger.ac.uk ----- Project Information Center project name: bz80B24 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator; 100% of reads Consensus quality: 117499 bases at least Q40 Consensus quality: 117570 bases at least Q30 Consensus quality: 117601 bases at least Q20 Insert size: 142059; sum-of-contigs Insert size: 125803; 4.8% error; agarose-fp Quality coverage: 10.26x in Q20 bases; sum-of-contigs Quality coverage: 11.70x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 9084: contig of 9084 bp in length * 9085 9184: gap of 100 bp * 9185 81290: contig of 72106 bp in length * 81291 81390: gap of 100 bp * 81391 105790: contig of 24400 bp in length * 105791 105890: gap of 100 bp * 105891 142359: contig of 36469 bp in length. Location/Qualifiers 1. .142359 /organism="Danio rerio" /mol_type="genomic DNA" /db_xref="taxon:7955" /clone="RP71-80B24" /clone_lib="RPCI-71" 1. .9084 /note="assembly fragment:01685 fragment_chain:1" 9185. .81290 /note="assembly fragment:02193 fragment_chain:1" 81391. .105790 /note="assembly fragment:02272"		

Gene	<50630. >50700
tRNA	/locus_tag="CHR10v09222002.tRNA-Gly-2"
Gene	50630. _50700
Gene	/locus_tag="CHR10v09222002.tRNA-Gly-2"
Gene	/product="tRNA-Gly"
Gene	complement(<51371..>65531)
mRNA	/locus_tag="OSUNBa0027L23.1"
mRNA	/note="contains Pfam profile PF00226 (DnaJ domain)"
mRNA	complement(join(<51371..51598,51735..51803,52273..52396,52749..52873,53371..53832,53913..54256,54617..54962,55048..55497,56316..56607,56727..57071,57167..57390,57628..58119,58220..58612,59157..59297,59401..59763,60050..62239,62854..62916,62990..63091,63228..63583,63882..63969,64073..64306,65064..>65531))
CDS	/locus_tag="OSUNBa0027L23.1"
CDS	complement(join(51371..51598,51735..51803,52273..52396,52749..52873,53371..53832,53913..54256,54617..54962,55048..55497,56316..56607,56727..57071,57167..57390,57628..58119,58220..58612,59157..59297,59401..59763,60050..62239,62854..62916,62990..63091,63228..63583,63882..63969,64073..64306,65064..>65531))
CDS	/locus_tag="OSUNBa0027L23.1"
CDS	/codon_start=1
CDS	/product="putative DnaJ domain containing protein"
CDS	/protein_id="AAP55138.1"
CDS	/db_xref="GI:31433654"
CDS	/translation="MDFAASHAASPPAPDASSPSSPRASSSSSSAAAEPEYLA RFFVVKHNRGVRRLICASSGLVTLDPATLVNTSYDASVGFDRAPEGNATEFTL TLTKDARGFKALRFPSSPLRAGILTELRLRPVHPVLDPPVLHLRRTHWAPFLKV TSVTELEGGHSDLRWCLDFMDSPAILIGDSYGRKTAGGGFVLCPLVGRKSKA FMAAGSTNTLIISYLTKTANGVGVSHVDNSQSMATDFTAKRANEAVGAAETHG EWSVTRLRPAAGHTASIESLSLIGIGRGLGDSVSRQLVLTNTSLVERPENYEV DVGFPSPVLIOTIRKISITYEHASTFYTPCOINLLVIVVSDEAGHPWAIIVRPLSTV SALVRPAEPQAFPEFNDGCPHUVASTSRDLATVLDVLOQSOCAIPVRLITM PGRHIDPCGARVHISHTVMEASMHKILATVAKAEVSSDTVPGAKIRLWRI REFNACIPYGVPIVNEVEPVLMALISLLPATPQNLPADAPLPFPSPKAAATNGF VACRLRLTSRVSASHVMAFPVAVGRIMGLNRSEGVAAEAGLVAMLIGGGPGDTS MLTDTRGESHATYMAKSVLFQPVVPIVNLRLPLSPVLSLSIVLEAMLCPE HGETTCHAFVELLROVAGRLRLPALFAHPAESVRETSVIMRTIAERDAIAAESMR DAALDKGALLHLNAFFYPAGERDVSQALVADWSQPALDILSLRLPGLVAYL HTRSDSDQNDVEPLSRQRRLIQQRSRGKSNAPVEQCPMSNNNGEDFEGHTN GPGADVHORHNAVPTAYTPFGISIDPSQAVPHGFVFEAFYENNHQTAQFLQSHA YLVDNSNGDLANSALHDSFVQAVVNTVPGSGRLCNWYGFRAFSLDHNRAALI WNERTRQELREALQAEHSLDVEKERTDDIVPGSSVTEADSETLPRLSNIVFVSF SYPLSKVEYCGQYUURLLESGSNVRAODFLRPDPVAFRALYHFLCADIGITVD GAVPDLGSSDDKCDLGRLDGFGGGGSSVRELCSFAMAIVVEQHNKVGPPDGTAIH TVLLDRTDRLALNRHLLLLKALMDSLNVSVCLVGGCVLADMLTVHAESERTAI PLQSNLRIELALFMEPLKEMWYIDKQGVQPLEKDAIRLWSKSIDWTIKCWASGM DKWLRIELRALWALKVYVPLTPSQIDGALSLHMSASHSDLDADAGEIVTPTPR KRILSPRCPLHVAQMLTCEPSIVEAAASLLKAIIVTRNPKAMIRLYTGAFYFALAY PGSNLLSIAQLFSATHQAFHGEAAVSSSLPLAKRSVIGLGLPESLYLVLESGP SAPAAWSDSDTPEIITWTKWRAEHLIRQLQHLGDFPQKLAQHCHSLYDVAPWPPV TTPNLKDEWCHRYYLRLNLCDEIRFFNFIVEHVEFLQSLAMWREELTRRPMDLSEE

Query Match 74.4%; Score 18.6; DB 15; Length 300029;  
Best Local Similarity 84.0%; Pred. No. 5.4e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

misc_feature 105891..142359
              /note="assembly_fragment:02115
              clone_end:SP6
              vector_side:right"

ORIGIN
Query Match      73.6%; Score 18.4; DB 14; Length 142359;
Best Local Similarity 95.0%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCC 21
    ||||| ||||| ||||| |||||
Db 26886 AAAAAAGACACAGCGAGCCC 26905

RESULT 31
CR936300
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CR936300 158420 bp DNA linear HTG 10-AUG-2005
Danio rerio chromosome 22 clone CH211-63J24, WORKING DRAFT
SEQUENCE.
CR936300
HTG; HTGS PHASE2; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 158420)
McGuire S.
Direct Submission
Submitted (09-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
On Aug 12, 2005 this sequence version replaced gi:71891561.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC63J24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 158405 bases at least Q40
Consensus quality: 158417 bases at least Q30
Consensus quality: 158419 bases at least Q20
Insert size: 158420; sum-of-contigs
Insert size: 168521; 6.4% error; agarose-fp
Quality coverage: 8.66x in Q20 bases; sum-of-contigs Quality
coverage: 8.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..158420
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="22"
/clone="CH211-63J24"
/clone_lib="CHORI-211"
1..158420
/note="assembly_fragment:01444"

misc_feature
1..158420
/note="assembly_fragment:01444"

ORIGIN
Query Match      73.6%; Score 18.4; DB 14; Length 158420;
Best Local Similarity 95.0%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

misc_feature 105891..142359
              /note="assembly_fragment:02115
              clone_end:SP6
              vector_side:right"

ORIGIN
Query Match      73.6%; Score 18.4; DB 14; Length 142359;
Best Local Similarity 95.0%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCC 21
    ||||| ||||| ||||| |||||
Db 26886 AAAAAAGACACAGCGAGCCC 26905

RESULT 32
CUSCC/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CUSCC
Cucumis sativus mRNA for catalase, partial cds, clone CRR36-5'.
D63386
D63386.1 GI:1199477

Cucumis sativus (cucumber)
Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
Toyama,T., Teramoto,H., Takeba,G. and Tsuji,H.
Cytokinin induces a rapid decrease in the levels of mRNAs for
catalase, 3-hydroxy-3-methylglutaryl CoA reductase, lectin and
other unidentified proteins in etiolated cotyledons of cucumber
Plant Cell Physiol. 36 (7), 1349-1359 (1995)
8564304
Toyama,T., Teramoto,H., Takeba,G. and Tsuji,H.
Unpublished
3 (bases 1 to 252)
Direct Submission
Submitted (10-JUL-1995) Tomoko Toyama, Kyoto University, Faculty of
Science, Department of Botany; Kitashirakawa Oiwakekyo, Sakyo-ku,
Kyoto, Kyoto 606-01, Japan
[E-mail: a5186@esakura.kudpc.kyoto-u.ac.jp, Tel: 075-753-4147,
Fax: 075-701-3262]
Location/Qualifiers
1..252
/organism="Cucumis sativus"
/mol_type="mRNA"
/db_xref="taxon:3659"
/clone="CRR36-5"
/tissue_type="Cotyledon"
/clone_lib="H.Teramoto"
/dev_stage="Seedling"
<1..>252
/codon_start=3
/product="catalase"
/protein_id="BAA09702.1"
/db_xref="GI:1199478"
/translation="ERVVHARGASAKGFLRVTHDITNLSCADFLRAPGVQTPVIVRFS
TVIHERGSPETLRDPRGPAVKFTREVRLVVVAFICT"

CDS
Query Match      72.8%; Score 18.2; DB 15; Length 252;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
Query Match      72.8%; Score 18.2; DB 15; Length 252;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAACGACACAGCGAGCCC 25
    ||||| ||||| ||||| |||||
Db 201 AAAACTTCACAGCGAGCCC 179

RESULT 33
HSA323695
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HSA323695
Homo sapiens genomic sequence surrounding NotI site, clone
NF3-161S.
AJ323695
AJ323695.1 GI:15868074

Homo sapiens (human)
Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

## REFERENCE

## AUTHORS

Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,  
Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V.,  
Levitsky, V.G., Kolchanov, N.A., Protodopov, A.I., Kashuba, V.I.,  
Kisselev, L.L., Wasserman, W., Wahlestedt, C., and Zabarovsky, E.R.

Noti flanking sequences: a tool for gene discovery and verification  
of the human genome

Nucleic Acids Res. 30 (14), 3163-3170 (2002)

## JOURNAL

## PUBMED

12136098

## REFERENCE

## AUTHORS

Zabarovsky, E.R.

Direct Submission

Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,

Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,

Sweden

## FEATURES

## source

Location/Qualifiers

1..1181

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="NR3-161S"

## ORIGIN

Query Match 72.8%; Score 18.2; DB 8; Length 1181;

Best Local Similarity 87.0%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACGACGAGCGGCCTC 23

||||| ||||| ||||| ||||| |||||

Db 444 AAAAAACGACGAGCGGCCTC 466

## RESULT 34

## AY374526

## LOCUS

AY374526 1824 bp RNA linear VRL 24-JAN-2005

Borna disease virus isolate H3515 N protein, X protein, and P

protein genes, complete cds.

## ACCESSION

AY374526

AY374526.1 GI:40018881

## KEYWORDS

Borna disease virus

Borna disease virus

Viruses; ssRNA negative-strand viruses; Mononegavirales;

Bornaviridae; Bornavirus.

1 (bases 1 to 1824)

Kolodziejek, J., Durrwald, R., Herzog, S., Ehrensperger, F., Lussy, H.

and Nowotny, N.

Genetic clustering of Borna disease virus natural animal isolates,

laboratory and vaccine strains strongly reflects their regional

geographical origin

J. Gen. Virol. 86 (PT 2), 385-398 (2005)

15659758

2 (bases 1 to 1824)

Kolodziejek, J., Durrwald, R., Herzog, S., Ehrensperger, F., Lussy, H.

and Nowotny, N.

Direct Submission

Submitted (27-AUG-2003) University of Veterinary Medicine, Vienna,

Clinical Virology Group, Institute of Virology, Veterinärplatz 1,

Vienna 1210, Austria

Location/Qualifiers

1..1824

/organism="Borna disease virus"

/mol\_type="genomic RNA"

/isolate="H3515"

/isolation\_source="brain"

/specific\_host="horse"

/db\_xref="taxon:12455"

/country="Germany; Bavaria"

/note="isolated in 1997"

1..1113

CDS

/note="p40; ORF I"  
/codon\_start=1  
/product="N protein"  
/protein\_id="AAR36933.1"  
/db\_xref="GI:40018882"  
/translation="MPPKRRLLVDDADAMEDQDLYEPPASLPKLPKFLQYTVGGSDPH  
PGIHEKDRIQNAVAKLQDQSRDMFHTVTPSLFCLLIPGLHAAFGVGVPPRESYLS  
TPVTRGEOTVVAKFYGEKTTQDLEISIFSHCCSLIIGVIGSSKIKAGAE  
OIKKRFKTMALARPSPHGETATLLOMENPHEADLWINGOPWGSFVLSLITTDPEP  
GKEMFDQIKLVASVQAQMTYTTIKYLAECMDATLTIPVAVEIRDFLEVSAKLEE  
ADLPFLGAIHRPAIKAPSPFNLASAAFTYWSKKENPTWAGYTRASTIQPGASVKET  
QLARYRREISRGEDGAELSGEISAIMRMIGVTGLN"

## CDS

1170..1433

/note="p10"

/codon\_start=1

/product="X protein"

/protein\_id="AAR36934.1"

/db\_xref="GI:40018883"

/translation="MSSDLRLTLLELVRRLNGSATIESGRLPFGGRRRSPDTTGTIGV  
TKTTEDPKECTDPTSRPAPEGPQEPLHDLRPRPANKGAAVE"

1219..1824

/note="p24; ORF II"

/codon\_start=1

/product="P protein"

/protein\_id="AAR36935.1"

/db\_xref="GI:40018884"

/translation="MAARPSSLVDSLEDEDPOTLRERSGSRPRKIPRNALTOPVD  
QLKDLRKNPSMISDPDQRTQREQLSNDLKLKLVTELAENSMTEAEVRGTGLDISA  
RIEAGFELSALQVETIQTQACDHSDSIRILGENIKILDRSMKMTMETKLMMEKVD  
LLYASTAVGTSAPMLPSHPAPRIYQPLPSAPTADWDIIP"

Query Match 72.8%; Score 18.2; DB 13; Length 1824;

Best Local Similarity 87.0%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## ORIGIN

QY 1 AAAAAACGACGAGCGGCCTC 23

||||| ||||| ||||| ||||| |||||

Db 362 AAAAAACGACGAGCGGCCTC 384

## RESULT 35

## AY374527

## LOCUS

AY374527 1824 bp RNA linear VRL 24-JAN-2005

Borna disease virus isolate H3519 N protein, X protein, and P

protein genes, complete cds.

## ACCESSION

AY374527

AY374527.1 GI:40018885

## KEYWORDS

Borna disease virus

Borna disease virus

Viruses; ssRNA negative-strand viruses; Mononegavirales;

Bornaviridae; Bornavirus.

1 (bases 1 to 1824)

Kolodziejek, J., Durrwald, R., Herzog, S., Ehrensperger, F., Lussy, H.

and Nowotny, N.

Genetic clustering of Borna disease virus natural animal isolates,

laboratory and vaccine strains strongly reflects their regional

geographical origin

J. Gen. Virol. 86 (PT 2), 385-398 (2005)

15659758

2 (bases 1 to 1824)

Kolodziejek, J., Durrwald, R., Herzog, S., Ehrensperger, F., Lussy, H.

and Nowotny, N.

Direct Submission

Submitted (27-AUG-2003) University of Veterinary Medicine, Vienna,

Clinical Virology Group, Institute of Virology, Veterinärplatz 1,

Vienna 1210, Austria

Location/Qualifiers

1..1824

/organism="Borna disease virus"

/mol\_type="genomic RNA"

/isolate="H3519"

/isolation\_source="brain"

/specific\_host="horse"

/db\_xref="taxon:12455"

/country="Germany; Bavaria"

/note="isolated in 1997"

1..1113

CDS

/isolation\_source="brain"  
/specific\_host="horse"  
/db\_xref="taxon:12455"  
/country="Germany; Bavaria"  
/note="isolated in 1997"  
1. .1113  
/note="p40; ORF I"  
/codon\_start=1  
/product="N protein"  
/protein\_id="AAR36936.1"  
/db\_xref="GI:40018886"  
/translation="MPPKRLVDADAMEDDQLYERPASLPKLPGLQVTVGGSDPH  
PGIGHEKDRQNAVALIDQSRDMFHTVPSLVFLCLLPLGLHAAPVHGVPRESYLS  
TPVTRGQTVTKTAKFYGEKTDLTLELSSIFSGCCSLIGVIGSSKIKAGAE  
OIKRFTMAALNRSHGFTATLQMFNEHALDWMNGOPWVGSVLILTTDPSF  
GKFMQIKLVASYAQMTYTTIKYELACMDATLIPVVAEIRDFEVSXKKEH  
ADUFPFGLAIRHFDALAPRSPFNLASAFYWSKKENPTMAGYRASTIQPGASVKET  
QLARYRRIRSGEDGAEUSGEISATMKMIGVTGLN"  
1170. .1433  
/note="p10"  
/codon\_start=1  
/product="X protein"  
/protein\_id="AAR36937.1"  
/db\_xref="GI:40018887"  
/translation="MSDDLRLTLLELVRRLNGSATIESGRLPGGRRSPDTTGTIGV  
TKTEDPKECTDPTSRPAPEGPQEPLDLRPRANKGAAVE"  
1219. .1824  
/note="p24; ORF II"  
/codon\_start=1  
/product="P protein"  
/protein\_id="AAR36938.1"  
/db\_xref="GI:40018888"  
/translation="MAAPSSLVDSLEEDDPOTLRERSGSRPRKIPRNALTOPVD  
QLKDKRNSMISDPQRTGRQLSNDLILKLVTELAENSMEAEVYRGTLGDLSA  
RISAGFESALQVETIQTAQRCHDSIRLIGENIKILDRSMKTMETMKLMEKVD  
LLYASTAVGTSAPMLPSHPAPPRIYPQLPSAPTADWDIIP"

CDS

CDS

CDS

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 1824;  
Best Local Similarity 87.0%; Pred. No. 1.1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAGACGACACGACGACCTC 23  
|||||  
Db 362 AAAAGACGACACGACGACCTC 384

RESULT 36  
AE011319/c  
LOCUS  
DEFINITION  
Leptospira interrogans serovar lai str. 56601 chromosome I, section  
128 of 397 of the complete sequence.  
ACCESSION  
AE011319 AE010300  
VERSION  
AE011319.1 GI:24194975  
KEYWORDS  
ORGANISM  
Leptospira interrogans serovar lai str. 56601  
Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
Xiong H., Lu, G., Lu, L. F., Jiang H. Q., Jia J., Tu, Y. F., Zhang, Y. X.,  
Gu, W. Y., Zhang, Y. Q., Cai, Z., Sheng, H. H., Yin, H. F., Zhang, Y.,  
Zhu, G. F., Wan, M., Huang, H. L., Qian, Z., Wang, S. Y., Ma, W., Yao, Z. J.,  
Shen, Y., Qiang, B. Q., Xia, Q. C., Guo, X. K., Danchin, A., Saint  
Giron, I., Somerville, R. L., Wen, Y. M., Shi, M. H., Chen, Z., Xu, J. G.  
and Zhao, G. P.

Unique physiological and pathogenic features of Leptospira  
interrogans revealed by whole-genome sequencing  
Nature 422 (6934), 888-893 (2003)  
12712204  
2 (bases 1 to 10029)  
Ren, S., Fu, G., Jiang, X., Zeng, R., Xiong, H., Lu, L., Lu, G., Jiang, H.,  
Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y.,

Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and  
Zhao, G.  
Direct Submission  
Submitted (12-MAR-2002) Chinese National Human Genome Center at  
Shanghai, 250 Bi Bo Road, Shanghai 201203, China  
Updated information will be available at our World Wide Web site  
(http://www.chgc.sh.cn/lep/). Comments to the authors are  
appreciated.  
FEATURES  
source  
1. .10029  
/organism="Leptospira interrogans serovar lai str. 56601"  
/mol\_type="genomic DNA"  
/strain="56601"  
/aerovar="Lai"  
/db\_xref="taxon:189518"  
/chromosome="I"  
/note="Strain 56601 is maintained by the Institute of  
Epidemiology and Microbiology, Chinese Academy of  
Preventive Medicine"  
/complement(179. .550)  
/locus\_tag="LA1393"  
/complement(179. .550)  
/locus\_tag="LA1393"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAN48592.1"  
/db\_xref="GI:24194976"  
/translation="MKSTFYFLFFGFLTNLIWYGYSSMELKNDYHILTESCEYSN  
LKLREDPFKGMTISFTKICNRDNLSECTSSSSKKKIQVQSVIPSPCHGRPDY  
GYFRFRNERLVRKVEAGYPC"  
616. .822  
/locus\_tag="LA1394"  
616. .822  
/locus\_tag="LA1394"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAN48593.1"  
/db\_xref="GI:24194977"  
/translation="MQKCVGTPADLEFYSLILKCGNSYRVELKDNLKLCKDEICR  
NSHKGCNKPFDFGTANLVVTVRF"  
complement(873. .1040)  
/locus\_tag="LA1395"  
complement(873. .1040)  
/locus\_tag="LA1395"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAN48594.1"  
/db\_xref="GI:24194978"  
/translation="MLNFPYFWVEAVRFGKFFSIRKLYFLQVKNLTLVGTLETKYLN  
IMQNFDRKF"  
complement(1155. .1733)  
/locus\_tag="LA1396"  
complement(1155. .1733)  
/locus\_tag="LA1396"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAN48595.1"  
/db\_xref="GI:24194979"  
/translation="MLKYSILRPSIIFITMNCVTKVYKGTWEKDLNTEFKRIVSVVP  
SSEVGNPKQLAAKTENYLSHHKFIYFPFSGNGICSSDKKQVQAIQOLKIREKQ  
TSDKLSALGKVLKCKGKGLVLAENSYRKNTEENQSLINTYTYLYKEITSKYN  
PYFYLQSLDKLDSPLTBEEXEKIEVEAR"  
complement(1747. .4701)  
/locus\_tag="LA1397"  
complement(1747. .4701)  
/locus\_tag="LA1397"  
/note="confirmed by proteomics"  
/codon\_start=1  
/transl\_table=11







Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczky,K.R.,  
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,  
Yan,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,  
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,  
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.  
Direct Submission  
Unpublished

## TITLE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Submitted (03-APR-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: KZU

Center clone name: CH230-unknown

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 35269 bases at least Q40

Consensus quality: 38450 bases at least Q30

Consensus quality: 41455 bases at least Q20

Estimated insert size: 25179; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 23 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1121: contig of 1121 bp in length

\* 1122: gap of unknown length

\* 1222: contig of 1051 bp in length

\* 2273: gap of unknown length

\* 3749: contig of 1377 bp in length

\* 3849: gap of unknown length

\* 3850: contig of 1594 bp in length

\* 5443: gap of unknown length

\* 5444: contig of 1059 bp in length

\* 5542: gap of unknown length

\* 6603: gap of unknown length

\* 7794: contig of 1092 bp in length

\* 7894: gap of unknown length

\* 7895: contig of 1264 bp in length

\* 9159: gap of unknown length

\* 9259: contig of 1582 bp in length

\* 10841: gap of unknown length

\* 10941: gap of unknown length

\* 12299: contig of 1358 bp in length

\* 12399: gap of unknown length

\* 13572: contig of 1173 bp in length

\* 13572: gap of unknown length

\* 13672: contig of 1124 bp in length

\* 14796: gap of unknown length

\* 14896: gap of unknown length

\* 16630: contig of 1734 bp in length

\* 16730: gap of unknown length

\* 17661: contig of 1031 bp in length

\* 17861: gap of unknown length

\* 18883: contig of 1022 bp in length

\* 18883: gap of unknown length

\* 20285: contig of 1302 bp in length

\* 20385: gap of unknown length

\* 20385: contig of 1800 bp in length

\* 22185: gap of unknown length

\* 22285 23970: contig of 1686 bp in length

\* 23971 24070: gap of unknown length

\* 24071 25716: contig of 1646 bp in length

\* 25717 25816: gap of unknown length

\* 25817 27932: contig of 2116 bp in length

\* 27933 28032: gap of unknown length

\* 30463 30562: contig of 2430 bp in length

\* 30563 31721: contig of 1159 bp in length

\* 31722 31821: gap of unknown length

\* 31822 34773: contig of 2852 bp in length

\* 34674 37857: gap of unknown length

\* 34774 37857: contig of 3084 bp in length.

## FEATURES

## source

1. .37857

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-unknown"

1122. .1221

/estimated\_length=unknown

2273. .2372

/estimated\_length=unknown

3750. .3849

/estimated\_length=unknown

5444. .5543

/estimated\_length=unknown

6603. .6702

/estimated\_length=unknown

7795. .7894

/estimated\_length=unknown

9159. .9258

/estimated\_length=unknown

10841. 10940

/estimated\_length=unknown

12299. .12398

/estimated\_length=unknown

13572. .13671

/estimated\_length=unknown

14796. 14895

/estimated\_length=unknown

16630. .16729

/estimated\_length=unknown

17761. .17860

/estimated\_length=unknown

18883. 18982

/estimated\_length=unknown

20285. .20384

/estimated\_length=unknown

22185. .22284

/estimated\_length=unknown

23971. .24070

/estimated\_length=unknown

25717. .25816

/estimated\_length=unknown

27933. .28032

/estimated\_length=unknown

30463. .30562

/estimated\_length=unknown

31722. .31821

/estimated\_length=unknown

34674. .34773

/estimated\_length=unknown

## ORIGIN

Query Match 72.8%; Score 18.2; DB 14; Length 37857;  
Best Local Similarity 87.0%; Pred. No. 9.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTC 23

|||||

Db 16600 AAAAAACATACAGCGAGCCCTC 16578

## RESULT 38

AC101757/c

LOCUS

DEFINITION

AC101757

VERSION

KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC101757 148875 bp DNA linear ROD 26-MAY-2004  
Mus musculus chromosome 8, clone RP24-545I11, complete sequence.

AC101757 148875 bp DNA linear ROD 26-MAY-2004  
Mus musculus chromosome 8, clone RP24-545I11, complete sequence.

AC101757.12 GI:47679264  
HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 148875)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus chromosome 8, clone RP24-545I11

Unpublished

2 (bases 1 to 148875)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,

Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,

Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,

Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

Jones,C., Kamat,A., Karatas,A., Kells,C., Lakoczek,K.,

Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrim,J.,

Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Ratta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 148875)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,

Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,

Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,

DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,

Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,

Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,

Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,

Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,

Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,

MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,

McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,

Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,

O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,

Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,

Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,

Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,

Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (21-APR-2004) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 148875)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,

Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J.,

Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,

DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,

Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,

Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,

Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,

Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,

Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,

DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.

Direct Submission

Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 26, 2004 this sequence version replaced gi:46195628.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIGR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submission@broad.mit.edu

----- Project Information

Center project name: Li7350

Center clone name: 545\_I\_11

-----

Location/Qualifiers

1..148875

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="8"

/map="8"

/clone="RP24-545I11"

/clone\_lib="RPCI-24 Male Mouse BAC"

1..4

/note="clone boundary"

clone end:SP6

site:MBot

repeat\_region

3564..3684

/rpt\_family="PB1D9"

repeat\_region

3685..3823

/rpt\_family="GA-rich"

repeat\_region

3826..4005

/rpt\_family="(GGAA)n"

repeat\_region

6548..6568

/rpt\_family="(T)n"

repeat\_region

7260..7305

/rpt\_family="(CA)n"

repeat\_region

7920..7940

/rpt\_family="AT-rich"

repeat\_region

8148..8211

/rpt\_family="(TCC)n"

unsure

complement(8172..8228)

/note="single clone coverage"

repeat\_region

8212..8258

/rpt\_family="(TC)n"

repeat\_region

8259..8290

/rpt\_family="(CA)n"

repeat\_region

8370..8394

/rpt\_family="AT-rich"

repeat\_region

complement(9340..9464)

/rpt\_family="PB1D7"

repeat\_region

complement(9690..9823)

/rpt\_family="B1F"

repeat\_region

13451..13471



in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII				EcoRI				HindIII			
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3570	3516	-----	8696	-----	8685	-----	882	-----	903	-----	228
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	5485
2067	2056	-----	6	-----	<800	-----	6382	-----	6532	-----	1112
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	4200
5793	5871	-----	2008	-----	1921	-----	512	-----	<800	-----	1944
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	3692
315	<800	-----	111	-----	<800	-----	449	-----	<800	-----	11463
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	906
3082	3132	-----	502	-----	<800	-----	1879	-----	1889	-----	5989
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	259
3061	3024	-----	2208	-----	2216	-----	3776	-----	3777	-----	2967
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	3235
1275	1223	-----	3870	-----	3784	-----	6154	-----	6139	-----	211
-----	-----	-----	-----	-----	-----	-----	-----	-----	<800	-----	2584
2532	2518	-----	2566	-----	2584	-----	690	-----	<800	-----	1051
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	6170
7238	7231	-----	100	-----	<800	-----	6	-----	<800	-----	2225
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	4526
4907	4698	-----	214	-----	<800	-----	2845	-----	2888	-----	910
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1051
325	<800	-----	958	-----	974	-----	169	-----	<800	-----	1902
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1461
3292	3365	-----	8629	-----	8685	-----	2853	-----	2888	-----	886
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	5674
745	<800	-----	2703	-----	2584	-----	1485	-----	1434	-----	5021
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	5112
1624	1561	-----	3028	-----	2967	-----	1278	-----	1255	-----	4385
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2451
2113	2125	-----	1125	-----	1112	-----	2124	-----	2098	-----	1442
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	12899
591	<800	-----	2174	-----	2216	-----	524	-----	<800	-----	4605
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	4623
8685	8655	-----	5329	-----	5296	-----	533	-----	<800	-----	2078
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	320
3668	3631	-----	2974	-----	2967	-----	3371	-----	3323	-----	509
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1588
1840	1838	-----	5981	-----	5989	-----	1889	-----	1889	-----	1576
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
660	<800	-----	8695	-----	8685	-----	459	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8518	8655	-----	1757	-----	1743	-----	261	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1934	1838	-----	2304	-----	2414	-----	2861	-----	2888	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5920	5871	-----	1045	-----	1051	-----	206	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3447	3365	-----	8191	-----	8148	-----	574	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1866	1838	-----	533	-----	<800	-----	1386	-----	1327	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3570	3516	-----	2471	-----	2584	-----	508	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3674	3631	-----	372	-----	<800	-----	862	-----	903	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4060	4023	-----	1002	-----	974	-----	1866	-----	1889	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
798	767	-----	6502	-----	6455	-----	129	-----	<800	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3294	3246	-----	6486	-----	6455	-----	8466	-----	8529	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4436	4385	-----	3244	-----	3200	-----	5025	-----	4978	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4692	4698	-----	3497	-----	3522	-----	2053	-----	2098	-----	-----

Query Match 72.8%; Score 18.2; DB 8; Length 150313;  
Best Local Similarity 87.0%; Pred. No. 8.7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAACGACACAGCGAGCCCTCGG 25  
DB 96576 AAAACGACACAGCGAGCCCTTGG 96598

RESULT 40  
AC099543/c

LOCUS 156107 bp DNA linear PRI 19-APR-2002  
DEFINITION Homo sapiens chromosome 3 clone RP11-554B20, complete sequence.

ACCESSION AC099543 AC055718

VERSION AC099543.2 GI:20198581

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 156107)

Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
 Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
 Direct Submission  
 2 (bases 1 to 156107)  
 Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
 Direct Submission  
 Submitted (15-NOV-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 156107)  
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
 Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
 Direct Submission  
 Submitted (19-APR-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 On Apr 19, 2002 this sequence version replaced gi:16930990.  
 -----  
 Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: uwgchgs@u.washington.edu  
 Drafting Center: BCM  
 -----  
 Project Information  
 Center project name: chr-3  
 Center clone name: RP11-S54B20 (bc0471)  
 -----  
 Summary Statistics  
 Sequencing vector: plasmid; 41% of reads  
 Sequencing vector: M13; L08021; 47% of reads  
 Sequencing vector: plasmid; L08752; 12% of reads  
 Chemistry: Dye-terminator ET; 30% of reads  
 Chemistry: Dye-terminator Big Dye; 69% of reads  
 Chemistry: Dye-terminator Big Dye; 69% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 156027 bases at least Q40  
 Consensus quality: 156104 bases at least Q30  
 Consensus quality: 156107 bases at least Q20  
 Insert size: 156107; sum-of-contigs  
 Quality coverage: 9.4x in Q20 bases; sum-of-contigs  
 -----  
 Overlapping Sequences:  
 5': Mapping in progress  
 3': Mapping in progress  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an  
 alternate chemistry or covered by high quality data (i.e., Phred  
 quality > 30); an attempt was made to resolve all sequencing  
 problems, such as compressions and repeats; all regions were  
 covered by at least one plasmid subclone or more than one M13  
 subclone; and the assembly was confirmed by restriction digest.  
 -----  
 Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest  
 fingerprinting. Comparison of the experimentally derived digest  
 fragments with sequence-predicted fragments is given below.  
 The electronically-digested sequence consists of both insert and  
 vector, in order to accurately represent the entire circular BAC.  
 Small fragments below a variable cutoff (approximately 400-800 bp)  
 are not resolved in the fingerprint and hence do not appear  
 in the table. There are no significant remaining discrepancies  
 between the experimental and predicted values. Uniquely ordered  
 fragments are separated by dashed lines.  
 -----  
 HindIII EcoRI  
 BglII

2313	2280	3937	3851	485	<800
-----	-----	-----	-----	-----	-----
4071	4108	11994	11832	1109	1128
-----	-----	-----	-----	-----	-----
2428	2471	76	<800	3204	3187
-----	-----	-----	-----	-----	-----
456	<800	1050	1058	3927	3843
-----	-----	-----	-----	-----	-----
3917	3853	853	822	6093	6087
-----	-----	-----	-----	-----	-----
630	<800	631	<800	8295	8286
-----	-----	-----	-----	-----	-----
523	<800	86	<800	609	<800
-----	-----	-----	-----	-----	-----
2335	2280	795	<800	1272	1248
-----	-----	-----	-----	-----	-----
1502	1503	2329	2288	15004	14888
-----	-----	-----	-----	-----	-----
5632	5643	481	<800	7351	7265
-----	-----	-----	-----	-----	-----
1585	1503	6885	6690	2421	2373
-----	-----	-----	-----	-----	-----
2313	2280	176	<800	765	<800
-----	-----	-----	-----	-----	-----
5031	4846	599	<800	3021	3058
-----	-----	-----	-----	-----	-----
855	837	2845	2894	1535	1530
-----	-----	-----	-----	-----	-----
6115	6065	7642	7642	1066	1128
-----	-----	-----	-----	-----	-----
1993	2052	981	1058	7161	7265
-----	-----	-----	-----	-----	-----
4182	4108	7874	7642	2429	2373
-----	-----	-----	-----	-----	-----
2183	2160	460	<800	44	<800
-----	-----	-----	-----	-----	-----
2250	2280	1455	1430	4482	4409
-----	-----	-----	-----	-----	-----
374	<800	-----	-----	1444	1530
-----	-----	-----	-----	-----	-----
2177	2160	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----
6776	6632	-----	-----	-----	-----
Query Match 72.8%; Score 18.2; DB 8; Length 156107;					
Best Local Similarity 87.0%; Pred. No. 8.7e+02;					
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;					
/	3	AAAACGACACAGCGAGCCTCGG	25		
5	150111	AAAACGACACAGCGAGCCTTGG	150089		
AC083820 190122 bp DNA linear HTG 09-OCT-2002					
Rattus norvegicus clone RP32-290H10, *** SEQUENCING IN PROGRESS					
***, 10 unordered pieces.					
AC083820					
AC083820.20 GI:23269309					
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.					
Rattus norvegicus (Norway rat)					
Rattus norvegicus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;					
Sciurognathi; Muridae; Muridae; Murinae; Rattus.					
1 (bases 1 to 190122)					
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,					
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,					
Barbak, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,					
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,					
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,					
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,					

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleaveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.B., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, B., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaequez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 190122)  
Worley, K.C.

Direct Submission  
Submitted (01-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 190122)  
Worley, K.C.

Direct Submission  
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: TUAH  
Center clone name: RP32-290H10  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 165695 bases at least Q40  
Consensus quality: 167639 bases at least Q30  
Consensus quality: 168696 bases at least Q20  
Estimated insert size: 203777; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length









genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: RUGX  
Center clone name: CH230-1L14  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 197323 bases at least Q40  
Consensus quality: 201468 bases at least Q30  
Consensus quality: 204719 bases at least Q20  
Estimated insert size: 219777; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record as arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 245806: contig of 245806 bp in length  
\* 245807 245906: gap of unknown length  
\* 245907 247137: contig of 1231 bp in length.

FEATURES

source  
1. .247137  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-1L14"  
misc\_feature  
1. .1367  
/note="wgs\_contig"  
misc\_feature  
137559. .138691  
/note="wgs\_contig"  
misc\_feature  
152473. .154721  
/note="wgs\_contig"  
misc\_feature  
226086. .228364  
/note="wgs\_contig"  
misc\_feature  
228415. .229561  
/note="wgs\_contig"  
misc\_feature  
243501. .245806  
/note="wgs\_contig"  
gap  
245807. .245906  
/estimated\_length=unknown

ORIGIN

Query Match 72.8%; Score 18.2; DB 14; Length 247137;  
Best Local Similarity 87.0%; Pred. No. 8.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTC 23  
|||||  
Db 182443 AAAAAAGGCCACAGCGCCCTC 182421

RESULT 45

AC096330/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-82B5, WORKING DRAFT SEQUENCE, 2  
AC096330  
ACCESSION AC096330.7 GI:30521925  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

AC098268  
AC098268.4 GI:22855777  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 247137)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Alzbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barbata, J., Benton, J., Bina, K., Bina, K., Blankenbush, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douchwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gortell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Hume, J., Jackson, L.E., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, B., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtarian, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 247137)  
Worley, K.C.  
Direct Submission  
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 247137)  
Worley, K.C.

Direct Submission  
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 14, 2002 this sequence version replaced gi:21953869.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.

#### REFERENCE AUTHORS

1 (bases 1 to 249741)  
 Mullen, D., Marie, E., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angulano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sittler, C.D., Smajis, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

#### TITLE JOURNAL

2 (bases 1 to 249741)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

#### REFERENCE AUTHORS

3 (bases 1 to 249741)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

#### COMMENT

On May 10, 2003 this sequence version replaced gi:24818522. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GEYS  
 Center clone name: CH230-82B5  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 242419 bases at least Q40  
 Consensus quality: 243103 bases at least Q30  
 Consensus quality: 243892 bases at least Q20  
 Estimated insert size: 258836; sum-of-contigs estimation  
 Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 248130: contig of 248130 bp in length  
 248131 248230: gap of unknown length  
 248231 249741: contig of 1511 bp in length.

#### FEATURES source

1. 249741  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-82B5"

#### misc\_feature

1. 1700  
 /note="wgs\_end\_extension"  
 clone\_end:T7  
 5115..6713  
 /note="wgs\_end\_extension"

#### misc\_feature

clone\_end:T7  
 7411..8207  
 /note="clone\_boundary"  
 clone\_end:T7

#### misc\_feature

end sequence: BHJ356129"  
 28181..246883  
 /note="clone boundary"  
 clone\_end:Sp6  
 site:EcoRI

#### misc\_feature

end sequence: BHJ356131"  
 248131..248230  
 /estimated\_length=unknown

#### ORIGIN

Query Match 72.8%; Score 18.2; DB 14; Length 249741;  
 Best Local Similarity 87.0%; Pred. No. 8.5e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGCAGCAGCGAGCCCTC 23

Db 185990 AAAAAAGCAGCAGCGAGCCCTC 185968

RESULT 46

```
AC080155/c
LOCUS       271995 bp      DNA      linear      HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone RP32-324M6, *** SEQUENCING IN PROGRESS ***,
12 unordered pieces.
AC080155
VERSION     AC080155.9  GI:23269316
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
REFERENCE   1 (bases 1 to 271995)
AUTHORS     Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, I.,
Barbata, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Burch, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, W.D., Bathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karleson, E., Kelly, S., Khan, J., King, J., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
Ogareny, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vaquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 271995)
Worley, K.C.
Direct Submission
Submitted (27-SEP-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 271995)
Worley, K.C.
Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 23, 2002 this sequence version replaced gi:22164800.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUA4
Center clone name: RP32-324M6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224636 bases at least Q40
Consensus quality: 228004 bases at least Q30
Consensus quality: 230671 bases at least Q20
Estimated insert size: 262923; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
table.
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
consists of 12 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 3147: contig of 3147 bp in length
* 3148 3247: gap of unknown length
* 3248 208352: contig of 205105 bp in length
* 208353 208452: gap of unknown length
* 208453 219789: contig of 11337 bp in length
* 219790 219889: gap of unknown length
* 219890 231262: contig of 11373 bp in length
* 231263 261189: contig of 29827 bp in length
* 261190 261289: gap of unknown length
* 261290 262472: contig of 1183 bp in length
* 262473 262572: gap of unknown length
* 262573 263781: contig of 1209 bp in length
* 263782 263881: gap of unknown length
* 263882 264967: contig of 1086 bp in length
* 264968 265067: gap of unknown length
* 265068 266917: gap of unknown length
* 266918 268484: contig of 1567 bp in length
* 268485 268584: gap of unknown length
* 268585 270311: contig of 1727 bp in length
* 270312 270411: gap of unknown length
* 270412 271995: contig of 1584 bp in length.
-----
Location/Qualifiers
1..271995
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="RP32-324M6"
1..1252
/contig="wgs_contig"
3148..3247
/estimated_length=unknown
180275..182104
/contig="wgs_contig"
182155..184160
/contig="wgs_contig"
188211..189659
/contig="wgs_contig"
FEATURES
source
misc_feature
gap
misc_feature
misc_feature
misc_feature
misc_feature
```





```

FGLTSGDAPFFYRNLWSTEGITGLGLRTLRYGVKQDRFAGKAMGWGNIELRWKFFD
FNIAGOHFALNLPFPMDFGRVWDEHNVLGDKDYKYISGLRGFRFIAMNQSTILMLDYAVS
KEDKQVFNFNHF"
4726..5001
/locus_tag="LIC12255"
4726..5001
/locus_tag="LIC12255"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA570827.1"
/db_xref="GI:45601346"
/translacion="MIFORISEIDIEILLFTTLGMIFITRIKIORKAIFTNLGART
LSLNGYFNYSIMIHSSSESDPNFIDRFLKCGNYHKLRFVELILKL"
complement (5668..6327)
/locus_tag="LIC12256"
complement (5668..6327)
/locus_tag="LIC12256"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="membrane metalloendopeptidase"
/protein_id="AA570828.1"
/db_xref="GI:45601347"
/translacion="WRSDFTVIKTKIRKILGYLIVQSLVFTFVLCSENKLEIPKIFE
SRRELGPDSIDKFTSEHNEYVGDGDFPVGPKNAGYMDQPGFNHGLGDMNAGV
RNDYGDVYVSHGLVKFVGDSGPGGTVMITLHPNGKRINSIYAHLSKLNVSXGD
QIRKGMIGRIGDANGRYGPHLHFEMRDFLPTGPGYSDPSGVLYNPKVIRKXKL
K"
6971..7483
/locus_tag="LIC12257"
6971..7483
/locus_tag="LIC12257"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA570829.1"
/db_xref="GI:45601348"
/translacion="MPKSKTNKKFOIANLLPEHREASIELVNQFRLINSIQLDGI
KIRPRAGTMDIVLKGTKGVLLGSGFLGDELVSLLIARTDEKPYLEEQKTLFIDL
AVTKRKQKSGYMKPLVLACESWAKQDEPKSIELAIVENENAVSFWKAMGYDPFYIR
FRKSPKWDSI"
complement (8468..9919)
/locus_tag="LIC12258"
complement (8468..9919)
/locus_tag="LIC12258"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA570830.1"
/db_xref="GI:45601349"
/translacion="MNRIFRNWNWIGLWILNSLPPIFGEGRLNLPPEARIPLDGKKL
EPSELAKEGWVYTAIPLLSSDPVRGQGGIRASLPNGKKTDLVYEYAVRSKLTLL
QLFTQVGNHFIQDPSFYILNTAFRWKSSLSLDYNPNQYVPGIGESSLSLSYS9PR
NLSGIRGVNANPDAYETSQYIRPSTGSEITPTVSDQYQYLNFTFFNGIDYT
FKAMKWVANEISRNIIIGHSIDGWHPSKDPYPAGSIWETPVNGESKLTEDYRAGKI
RGVNGGDIYVFRAGIYDTRDPEPDPRGILAEINVANYSKRTGSDFNKNIFFPOTKY
FYKILPDVEELVATRVALGYTSSGAPFSEVRYMWSLDQPMGTGIGLOTMRGYRODR
FVAPVVGSAEPRWRPAFKIFDELFTLSLVPFVDVGRWDSSEKINLQGYKHSWS
GFRILWQATVILIDFAKSKDSQMFVDFSHAF"
10798..14472
/locus_tag="LIC12259"
10798..14472
/locus_tag="LIC12259"
/notes="contains alpha integrin like repeat domains;
identified by sequence similarity; putative; ORF located
```

```

using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="cytoplasmic membrane protein"
/protein_id="AA570831.1"
/db_xref="GI:45601350"
/translacion="MISVITKTFISFSLIFGGIGCVQCKGNSMPFFAYLDFSNRSADS
SPNVSOISPGPNVTGSLNTSIQVFSQKLDSSSIQSIOQLTOGNTIIPGNFTSTEK
TLLNPNRSSLAASYVSYSISKDLSKMDGSSLSRDYTWSTNTTIDVLAADVSLRTP
TIGANLVNNTSVQIAFETWNCISINIVNFKNNVTNVLSPNSVCLGSAVILTPN
NPLAFNTYRVYDILSTAKDLANNPLNAYNMTFTTGVAADPDTVTVSVNTPNAQNV
PINETISIAFSEPNCLATIIIGSIVLDDNIIIPGSVNGPGCTGTASTPLNLTPTNT
NITVTSNAITDLQNNPTPTSTGTTAAADVDTQPTVFTFVPSANANGVGTNANPMV
VSEPMSCASVTSASERLKROATPGPDLLTPPSAVVTPANAAGVPTNGGVSAFSEAM
DOGLDVTNPLIPINNFTTGPDPDTPPSAVVTPANAAGVPTNGGVSAFSEAM
SNNALNGNAYWSTFTGVAADPDPVQVSLGSGVAVATNANITVAFNETINCST
LNFTVNGINGYAVNCSSATFIPSALTPLNAGNTYATILTVNDIVNPIGAAFGWS
FTTGAAPDVTTPVVTIQNLRNNSIVESGFVGTATDAGTITSVEVSLDNGAFVPAATG
NPMFKLPDINTWKSQHTI IARAKDLANNLTAAISVRKNNKDIINGDYVDLV
SAEYGOGLLYIFHSSGNAGMTITNAQASAKIIVGAEEFGRTVSMGLDNGDGPADVI
SGAPWNGAQGRVYIFHSSGNAGVNI SPSGFATKTISANAGARFGDSIVTGDLDNGD
Query Match 72.8%; Score 18.2; DB 1; Length 301258;
Best Local Similarity 87.0%; Pred. No. 8.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACACACAGCAGCGCCCTC 23
|||||
Db 117718 AAAAAACGACAAAGCGATCGCTC 117740
RESULT 48
CR378680 349751 bp DNA linear BCT 17-APR-2005
LOCUS Photobacterium profundum SS9 chromosome 2; segment 6/7.
DEFINITION CR378680 CR354532
ACCESSION CR378680.1 GI:46916688
VERSION CR378680.1 GI:46916688
KEYWORDS complete genome.
SOURCE Photobacterium profundum SS9
ORGANISM Photobacterium profundum SS9
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Photobacterium.
REFERENCE 1
AUTHORS Vezzi, A., Campanaro, S., D'Angelo, M., Simonato, F., Vitulo, N.,
Lauro, F., Cesaro, A., Malacrida, G., Simonati, B., Camata, N.,
Bartlett, D. and Valie, G.
TITLE Genome Analysis of Photobacterium profundum reveals the complexity
of high pressure adaptations
JOURNAL Unpublished
AUTHORS 2
Cestaro, A.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2004) Cestaro A., CRIBI, Biotechnology Centre,
University of Padova via U. Bassi 58/B, 35131 Padova, Italy
FEATURES
source
1..349751
/organism="Photobacterium profundum SS9"
/mol_type="genomic DNA"
/strain="ss9"
/db_xref="taxon:298386"
/chromosome="2"
complement (314..1732)
/genes="STM1598"
/locus_tag="PBRB1592"
complement (314..1732)
/genes="STM1598"
/locus_tag="PBRB1592"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
```



```
/product="Hypothetical transcriptional regulator"
/protein_id="CAG23452.1"
/db_xref="GI:46916689"
/db_xref="GOA:Q6LGX8"
/db_xref="UniProt:IPR000524"
/db_xref="InterPro:IPR000524"
/db_xref="UniProt/TREMBL:O6LGX8"
/translations="MAPQQRHYAVATEIQOQIEQRIWLPFGDRIPSRVMTKKHSISP
MTVKAYELLETEGWTYAKPKSGVFTVAHLNRLEAQPQMPQVNNRSIKINAHIFDVL
TACKRPDVPFGSAPDPDLPPLKQLGRALAOQTKMTQAQSTVTDLPNGNIALRRRAIA
KRYRIDIVSINDIVITSGMESLGLSLMAVTPGDTVAIESPAFYGAQOAVLRRL
KAVEIATHGTGMSVALEAAIKTHDIKACMLMSKYONPLGASMPDENKAIANNLLMK
NIPTLEDNNELYSFSESQPTPIKAYDKRMWVLCGSKCLAPGRVGVVAGQYA
KQIEQLQGLTSTAVNQALAQYVHLGGYDHLRLRLKQARQQOQMLDAIEQYFP
REQITRPEGYFLWVLPKHINTSDLLQQLDVHNISITAPGSLFASDDRYKNCMRIN
CSYPMNERTQOALKTVANCIK"
complement(2013..3242)
/gene="VP0623"
/locus_tag="PBPRB1593"
complement(2013..3242)
/gene="VP0623"
/locus_tag="PBPRB1593"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical glycine/D-amino acid oxidase"
/protein_id="CAG23453.1"
/db_xref="GI:46916690"
/db_xref="GOA:Q6LGX7"
/db_xref="UniProt:IPR006076"
/db_xref="UniProt/TREMBL:O6LGX7"
/translations="MIGLSAWLAEGHVDVTIDRQSKAETSANAGQISVGYSS
FWAAGIPVAKLWAKHAPLKI KPSLSPLDVFWASKLANCQERYQNNKARMLRV
ANYSRDCLDLRKHQPTINEYQGGTQLQVFNAAQIAEAVAKDIQVLAESRTEYELDV
AGTAAEPGLNWKVGLVLRPDPDETGYCFQOQLTEIAKAGVKHFNVVDQKL
NOTQKVSYSTSGALKADAYVMAGSYSSLSPLGLSIPVYPVKYSLTIPVDE
TQAPSVIMDETYPMTDFQRIKRVAGTAELAGYDLDLSEQRKATIAMVQDLFNG
GDMSKARYWGLRPVTDGAPIGKTFPSNLFNTYTGHTLGTWACGSARLLADVLSG
VTDIPDPDGLSIARYNA"
/3409..4353
/gene="DSDC"
/locus_tag="PBPRB1594"
/3409..4353
/gene="DSDC"
/locus_tag="PBPRB1594"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical transcription regulator LyxR
family"
/protein_id="CAG23454.1"
/db_xref="GI:46916691"
/db_xref="GOA:Q6LGX6"
/db_xref="UniProt:IPR000847"
/db_xref="InterPro:IPR005119"
/db_xref="UniProt/TREMBL:O6LGX6"
/translations="MLAGLATFETAARFGSFTQAAEHLHITTAISQOISHLRQELS
VLPFRHSRGIHLDEGNQLYRVNQSLADIRDVINOIQKAHADGIEHLKLTSPFAFK
WLVRPHDFTQYDPDKIHTFAEGALVDHNDTDFDLAIDYQOTPIQTRNKSQISILIM
AEKLLPVNSDPYKGFNNPQDTSQNIKEKVLHLDHAMPWRHAARDKDEWCFKENG
LTPLNNTSRQGHYFNRTDMMAAAAGLVALARCAIVEEDFSKGNLVSPPPELDA
KAGYLIQHHSIPAIECFKAWLRQQVNRVYCGSGSGTCL"
/5180..5521
/locus_tag="PBPRB1595"
/5180..5521
/locus_tag="PBPRB1595"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental

gene
CDS

/transl_table=11
/product="Hypothetical protein"
/protein_id="CAG23455.1"
/db_xref="GI:46916692"
/db_xref="UniProt/TREMBL:O6LGX5"
/translations="MLKIYALITDLDLVHSHGISGLSKSYDHHAILKVKHFVSETDK
GIVFALVDENSKLHFKLSRSLFNVSRILEVENHRATCFYIIKSMKMTIETEFAD
DRKCLTETVRP"
/5810..6286
/locus_tag="PBPRB1596"
/5810..6286
/locus_tag="PBPRB1596"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAG23456.1"
/db_xref="GI:46916693"
/db_xref="UniProt/TREMBL:O6LGX4"
/translations="MIYKLLPGCIISALLTSTSLATEISONKHGHPLEYSININ
DSESTAYVTPDMQSDTYGDPILIGSDIQESTVADNPISIIYOTGHELVVTTITNTKI
NIQARVMKISPLDKTSVQTHIKTLAPDNTSPHTSCQOVFNGBEIIQSIQVCLASPS"
/6876..7454
/locus_tag="PBPRB1597"
/6876..7454
/locus_tag="PBPRB1597"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAG23457.1"
/db_xref="GI:46916694"
/db_xref="UniProt/TREMBL:O6LGX3"
/translations="MKPIASDNAQDDGLKHISMVHNI SAYHSLIILREKTKVLDDVL
SALHSAVARMTLPQITALTATTTTHATQORVYI VHSMEPEMSECGVSALWSSPLPAL
YCLFLAESEGHLSQTQDIVALLRVDFDKRPLDDTAIEIELYRARDHECALAADIL
QQPLLHKEGLIPAIRDCKPIGFESFEDLIARQ"
/7512..7847
/locus_tag="PBPRB1598"
/7512..7847
/locus_tag="PBPRB1598"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Hypothetical conjugal transfer protein, TrbC"
/protein_id="CAG23458.1"
/db_xref="GI:46916695"
/db_xref="InterPro:IPR007039"
/db_xref="UniProt/TREMBL:O6LGX2"
/translations="MRCQWRQWSQWLKANIAVFGFWPFLAQASNTGMPYESPLEK
VNSITGPVAFGASVLGIVAAIGLIFGGMNGFIQRLMILSLVISIVNATNLMSPFL
FNVSSTLV"
/7858..8151
/locus_tag="PBPRB1599"
/7858..8151
/locus_tag="PBPRB1599"
/notes="predicted by orpheus program
predicted by glimmer program"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical TrbD protein"
/protein_id="CAG23459.1"
/
```

gene /db\_xref="GI:46916696"  
/db\_xref="InterPro:IPR007909"  
/db\_xref="UniProt/TREMBL:Q6LGX1"  
/translation="WNTDEPLVCRFPVFNRRSHLFLGGERELVLSMVLICIVLIVALQ  
NIYTAMGVLLIVSITVARMMAKSDPOLTKTYRRFARYQRFYPSNGMKQYGG"  
8153. .10549  
/gene="AGR\_PTI "  
/locus\_tag="PBRB1600"  
8153. .10549  
/gene="AGR\_PTI "  
/locus\_tag="PBRB1600"  
/note="predicted by orpheus program  
predicted by glimmer program"  
/codon\_start=1  
/evidence=not\_experimental

CDS  
72.8%; Score 18.2; DB 1; Length 349751;  
Best Local Similarity 87.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTC 23  
Db 4946 AAAAAACTACACGAGCGCCGCG 4968

RESULT 49  
AX432478/c  
LOCUS AX432478 1368 bp DNA linear PAT 28-JUN-2002  
DEFINITION Sequence 893 from Patent WO0229113.  
ACCESSION AX432478  
VERSION AX432478.1 GI:21657282  
KEYWORDS  
SOURCE Bacillus licheniformis  
ORGANISM Bacillus licheniformis  
REFERENCE 1  
AUTHORS Berka,R. and Clausen,I.G.  
TITLE Methods for monitoring multiple gene expression  
JOURNAL Patent: WO 0229113-A 893 11-APR-2002;  
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)  
FEATURES  
source  
1. .1368  
/organism="Bacillus licheniformis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1402"

ORIGIN  
Query Match 71.2%; Score 17.8; DB 6; Length 1368;  
Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAACGACACGAGCGCCCT 22  
Db 1250 AAAAAACGACAAAGGAGCCCT 1230

RESULT 50  
AX954557/c  
LOCUS AX954557 2347 bp DNA linear PAT 08-JAN-2004  
DEFINITION Sequence 23 from Patent WO03093453.  
ACCESSION AX954557  
VERSION AX954557.1 GI:40783930  
KEYWORDS  
SOURCE Bacillus licheniformis  
ORGANISM Bacillus licheniformis  
REFERENCE 1  
AUTHORS Andersen,J.T., j Rgensen,S.T., Rasmussen,M.D., Olsen,P.B. and Clausen,I.G.  
TITLE Improved bacillus host cell  
JOURNAL Patent: WO 03093453-A 23 13-NOV-2003;  
Novozymes A/S (DK)  
FEATURES  
Location/Qualifiers

source  
1. .2347  
/organism="Bacillus licheniformis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1402"  
513. .1859  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAF05474.1"  
/db\_xref="GI:40783931"  
/translation="MISAAASLLLLTCFFPLPSTAQTAIGNETKQRLTYPVLTAKTP  
EAGFSSKKKAVDRLEIQDYKAGFPGAALILIKDGKLIKKEVGYKQKINGLIALKH  
PKQKANTMFDLASNTMYAVNFALQHLVSTGKLDLNKNI SQYLPDFKDHPEDDYKGG  
NRLVIDLLHNAGFPASWNYDPSAGHLYSQRSKLTLEYLVKTPLAYEPTKQIYS  
DIDYMLGLIIEKITNEELDTFVENRFYRPLGLRHTLFNPLQKGFKRSHFAATERLGN  
TRDGTISFPNIRTYYTLQGEVHDEKAFYSMEGVSAGLGFSTVDDAAVLQLQVMLNGGY  
GRQHLFSSVISQFTEPSKTNPTTYGLGWRNLNGNTDMEWMFGKHASSKAYGHTGTGTV  
TIIDFVYQIGIVLLTNKKHSPVINPKENPNQFEGDEFATGKYGSVITAYVEALHHQ"  
ORIGIN  
Query Match 71.2%; Score 17.8; DB 6; Length 2347;  
Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 AAAAAACGACACGAGCGCCCT 22  
Db 327 AAAAAACGACAAAGGAGCCCT 307  
Search completed: February 3, 2006, 21:26:41  
Job time : 877 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-15

Perfect score: 25

Sequence: 1 aaaaacgacacagcgagcctcgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21:\*

1: geneseqn1908s:\*

2: geneseqn1908s:\*

3: geneseqn2008s:\*

4: geneseqn2008s:\*

5: geneseqn2008s:\*

6: geneseqn2008s:\*

7: geneseqn2008s:\*

8: geneseqn2008s:\*

9: geneseqn2008s:\*

10: geneseqn2008s:\*

11: geneseqn2008s:\*

12: geneseqn2008s:\*

13: geneseqn2008s:\*

14: geneseqn2008s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.8	79.2	770	6	ABQ26883
2	19.8	79.2	770	6	ABQ26882
3	18.8	75.2	763	6	ABQ26882
4	18.6	74.4	378	8	ABQ26882
5	17.8	71.2	1368	6	ABQ26882
6	17.8	71.2	1368	6	ABQ26882
7	17.8	71.2	1368	6	ABQ26882
8	17.6	70.4	559	6	ABQ26882
9	17.6	70.4	559	6	ABQ26882
10	17.6	70.4	559	6	ABQ26882
11	17.6	70.4	559	6	ABQ26882
12	17.6	70.4	559	6	ABQ26882
13	17.6	70.4	559	6	ABQ26882
14	17.6	70.4	559	6	ABQ26882
15	17.6	70.4	559	6	ABQ26882
16	17.6	70.4	559	6	ABQ26882
17	17.6	70.4	559	6	ABQ26882
18	17.2	68.8	2000	11	ABQ26882
19	17.2	68.8	6240	10	ABQ26882

93	16.2	64.8	1061	6	ABQ21137	Abq21137 Oligonucleotide
94	16.2	64.8	1409	1	AA50474	Sequence
95	16.2	64.8	2013	3	AA229206	Rat myel
96	16.2	64.8	2013	3	AA229206	Rat myel
97	16.2	64.8	2185	10	ADD48351	Add48351 Rat gene
98	16.2	64.8	2426	6	ABQ70640	Listeria
99	16.2	64.8	2468	12	ADP72650	Renal tox
100	16.2	64.8	2468	13	ADP72650	Renal tox
101	16.2	64.8	2474	10	ABT41975	Toxicity
102	16.2	64.8	2475	9	ACF25356	Rat MAG g
103	16.2	64.8	2475	9	ACF25356	Rat MAG g
104	16.2	64.8	2475	12	ADP58278	Toxicity
105	16.2	64.8	2475	12	ADP58278	Toxicity
106	16.2	64.8	2475	12	ADP58278	Toxicity
107	16.2	64.8	2475	12	ADP58278	Toxicity
108	16.2	64.8	2475	12	ADP58278	Toxicity
109	16.2	64.8	2475	12	ADP58278	Toxicity
110	16.2	64.8	2475	12	ADP58278	Toxicity
111	16.2	64.8	2475	12	ADP58278	Toxicity
112	16.2	64.8	2475	12	ADP58278	Toxicity
113	16.2	64.8	2475	12	ADP58278	Toxicity
114	16.2	64.8	2475	12	ADP58278	Toxicity
115	16.2	64.8	2475	12	ADP58278	Toxicity
116	16.2	64.8	2475	12	ADP58278	Toxicity
117	16.2	64.8	2475	12	ADP58278	Toxicity
118	16.2	64.8	2475	12	ADP58278	Toxicity
119	16.2	64.8	2475	12	ADP58278	Toxicity
120	16.2	64.8	2475	12	ADP58278	Toxicity
121	16.2	64.8	2475	12	ADP58278	Toxicity
122	16.2	64.8	2475	12	ADP58278	Toxicity
123	16.2	64.8	2475	12	ADP58278	Toxicity
124	16.2	64.8	2475	12	ADP58278	Toxicity
125	16.2	64.8	2475	12	ADP58278	Toxicity
126	16.2	64.8	2475	12	ADP58278	Toxicity
127	16.2	64.8	2475	12	ADP58278	Toxicity
128	16.2	64.8	2475	12	ADP58278	Toxicity
129	16.2	64.8	2475	12	ADP58278	Toxicity
130	16.2	64.8	2475	12	ADP58278	Toxicity
131	16.2	64.8	2475	12	ADP58278	Toxicity
132	16.2	64.8	2475	12	ADP58278	Toxicity
133	16.2	64.8	2475	12	ADP58278	Toxicity
134	16.2	64.8	2475	12	ADP58278	Toxicity
135	16.2	64.8	2475	12	ADP58278	Toxicity
136	16.2	64.8	2475	12	ADP58278	Toxicity
137	16.2	64.8	2475	12	ADP58278	Toxicity
138	16.2	64.8	2475	12	ADP58278	Toxicity
139	16.2	64.8	2475	12	ADP58278	Toxicity
140	16.2	64.8	2475	12	ADP58278	Toxicity
141	16.2	64.8	2475	12	ADP58278	Toxicity
142	16.2	64.8	2475	12	ADP58278	Toxicity
143	16.2	64.8	2475	12	ADP58278	Toxicity
144	16.2	64.8	2475	12	ADP58278	Toxicity
145	16.2	64.8	2475	12	ADP58278	Toxicity
146	16.2	64.8	2475	12	ADP58278	Toxicity
147	16.2	64.8	2475	12	ADP58278	Toxicity
148	16.2	64.8	2475	12	ADP58278	Toxicity
149	16.2	64.8	2475	12	ADP58278	Toxicity
150	16.2	64.8	2475	12	ADP58278	Toxicity

## ALIGNMENTS

RESULT 1	
ABQ26883	
ID	ABQ26883 standard; DNA; 770 BP.
XX	
AC	ABQ26883;
XX	
DT	12-JUL-2002 (first entry)
XX	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 13474.
XX	

RESULT 2	
ABQ26882/c	
ID	ABQ26882 standard; DNA; 770 BP.
XX	
AC	ABQ26882;
XX	
DT	12-JUL-2002 (first entry)
XX	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 13473.
XX	
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
XX	SNP; cell differentiation; ds.
XX	
OS	Homo sapiens.
XX	
XX	WO200218632-A2.
XX	
XX	07-MAR-2002.
XX	
XX	01-SEP-2001; 2001WO-EP010074.
XX	
XX	01-SEP-2000; 2000DE-01043826.
XX	
XX	05-SEP-2000; 2000DE-01044543.
XX	
XX	(EPIG-) EPIGENOMICS AG.
XX	
XX	Olek A, Piepenbrock C, Berlin K, Guetig D;
XX	
XX	WPI; 2002-371829/40.
XX	
XX	Determining the degree of cytosine methylation in genomic DNA, useful for
XX	diagnosis and prognosis, comprises selective hybridization of amplicons
XX	from chemically treated DNA.
XX	
XX	Claim 12; 56pp + Sequence Listing; 56pp; German.
XX	
XX	This invention describes a novel method for determining the degree of
XX	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX	genomic sample of DNA. The sample is treated chemically to convert
XX	cytosine (C) but not methylated C, to uracil, then part of the genomic
XX	DNA that contains the target C is amplified to form a labeled amplicon.
XX	The amplicon is hybridised to two classes, each with at least one member,
XX	of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX	degree of hybridisation to both classes is determined from the label on
XX	the amplicon. From the ratio of labels hybridised to the two classes of
XX	oligomers, the degree of methylation is calculated. The method is used:
XX	(i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX	and of a wide range of diseases, e.g. cancer, disorders of the central
XX	nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX	particularly by detecting mutations or single nucleotide polymorphisms
XX	(SNP's); and (ii) for differentiation of cell or tissue types and for
XX	investigating cell differentiation. The method allows the methylation
XX	status of many C residues to be determined simultaneously. ABQ1410-
XX	ABQ54121 represent genomic DNA sequences used to illustrate the method
XX	for determining the degree of cytosine methylation described in the
XX	disclosure of the invention
XX	
XX	Sequence 770 BP; 258 A; 318 C; 85 G; 109 T; 0 U; 0 Other;
SQ	
Query Match	79.2%; Score 19.8; DB 6; Length 770;
Best Local Similarity	91.3%; Pred. No. 29;
Matches	21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 AAAAAACGACACGAGCCCTC 23
DB	362 AAAAAACGACACGAGCCCTC 384

KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX W0200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention

XX SQ Sequence 770 BP; 109 A; 85 C; 318 G; 258 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 770;

Best Local Similarity 91.3%; Pred. No. 29;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGCGGCCCTC 23

Db 409 AAAAAACGACACCGCGGCCCTC 387

RESULT 3

ABQ576772

ID ABS76772 standard; cDNA; 763 BP.

XX AC ABS76772;

XX DT 12-DEC-2002 (first entry)

XX DE Frog embryonic gene sequence Q9924949.

XX KW Frog; ss; embryonic development; developmental disorder; microarray;

XX KW cell differentiation.

XX OS Xenopus laevis.

XX PN US2002081610-A1.

XX 27-JUN-2002.

XX 23-JUL-2001; 2001US-00910943.

XX 21-JUL-2000; 2000US-0219658P.

XX (UYRQ ) UNIV ROCKEFELLER.

XX Hemmati-Brivanlou A, Altmann CR;

XX WPI; 2002-626534/67.

XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful  
PT to identify genes involved in embryonic development, to identify  
PT different types of embryonic cells, and to diagnose developmental  
PT disorders.

XX Claim 1; Page 42; 823pp; English.

XX The invention relates to a nucleic acid array, where each coordinate  
CC contains a single nucleic acid species having one of 770 nucleotide  
CC sequences (appearing as ABS76747-ABS77516) a of a xenopus embryonic gene  
CC product, or its complement or hybridisable fragment of not less than 20  
CC contiguous nucleotides of one of those sequences. Also included are  
CC detecting differential expression of embryonic genes, comprising: (a)  
CC contacting a nucleic acid array comprising genes expressed in embryonic  
CC but not mature cells with nucleic acids from sample and control cells;  
CC and (b) detecting differential hybridisation of nucleic acids from the  
CC sample cells relative to the control cells; and detecting defects in  
CC development, comprising: (a) contacting nucleic acids from test cells  
CC undergoing development with a nucleic acid array of gene products known  
CC to play a fundamental role in the development process; and (b) detecting  
CC a difference in expression of a fundamental gene in the sample cells  
CC relative to a standard. The invention is useful to identify genes  
CC involved in embryonic development and related processes such as cell  
CC differentiation. This would be useful for diagnosing developmental  
CC disorders and for identifying different types of embryonic cells. The  
CC present sequence is one of the 770 Xenopus embryonic cDNA sequences

XX SQ Sequence 763 BP; 181 A; 223 C; 208 G; 141 T; 0 U; 10 Other;

Query Match 75.2%; Score 18.8; DB 6; Length 763;

Best Local Similarity 90.9%; Pred. No. 84;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AAACGACACGACGCGCCCTCGG 25

Db 549 AAACGACACGACGACGCGCCCTCGG 570

RESULT 4

ABX96483/c

ID ABX96483 standard; cDNA; 378 BP.

XX AC ABX96483;

XX DT 13-MAY-2003 (first entry)

XX DE Rice endosperm expression sequence label #24.

XX KW Rice; endosperm expression sequence label; gene chip; babel technique;  
XX KW microarray; ss.

XX OS Oryza sp.

XX PN CN1366050-A.

XX PD 28-AUG-2002.

XX 31-OCT-2001; 2001CN-00135874.

XX 31-OCT-2001; 2001CN-00135874.

```
XX (UYZH-) UNIV ZHEJIANG.
XX
XX Dong H, Li D;
XX
XX WPI; 2003-240400/24.
XX
XX Rice endosperm expression sequence label and genechip prepared from it.
XX
XX Claim 1; Page 19 (Disclosure); 29pp; Chinese.
XX
XX The invention describes a novel rice endosperm expression sequence label
XX and the gene chip formed from it. The expression sequence label technique
XX is used to create a rice endosperm cDNA library. Non-redundant expression
XX sequence labels are then combined and used to create a gene chip by
XX microarray techniques. This sequence represents a rice endosperm
XX expression sequence
XX
XX Sequence 378 BP; 90 A; 119 C; 87 G; 82 T; 0 U; 0 Other;
SQ
Query Match 74.4%; Score 18.6; DB 8; Length 378;
Best Local Similarity 84.0%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
DB 235 AAAAAACGACACAGCGGGGCTCTGGG 211
RESULT 5
ABK73602/C
ID ABK73602 standard; DNA; 1368 BP.
XX
XX AC ABK73602;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Bacillus licheniformis genomic sequence tag (GST) #893.
XX
XX Differential gene expression; genomic sequenced tag; GST;
XX altered culture condition; environmental stress;
XX physiological provocation; ds.
XX
XX OS Bacillus licheniformis.
XX
XX PN WO200229113-A2.
XX
XX PD 11-APR-2002.
XX
XX PF 05-OCT-2001; 2001WO-US031437.
XX
XX PR 06-OCT-2000; 2000US-00680598.
XX
XX PR 27-MAR-2001; 2001US-0279526P.
XX
XX (NOVO ) NOVOZYMES BIOTECH INC.
XX (NOVO ) NOVOZYMES AS.
XX
XX Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.
XX
XX Monitoring differential expression of several genes in first Bacillus
XX cell relative to expression of same genes in one or more second Bacillus
XX cells, by using substrate containing Bacillus genomic sequenced tag
XX array.
XX
XX Claim 4; SEQ ID NO 893; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX
relative gene expression by an observed hybridisation reporter signal of
a spot in the array. The method is useful for measuring the expression of
genes in a first Bacillus cell relative to expression of the same genes
in one or more second Bacillus cells. The method is useful for monitoring
global expression of several genes from a Bacillus cell, discovering new
genes, identifying possible functions of unknown open reading frames and
monitoring gene copy number variation and stability. Monitoring changes
in expression of genes may be used to provide a representation of the way
in which Bacillus cells adapt to changes in culture conditions,
environmental stress or other physiological provocation. Extensive follow
up characterisation is unnecessary, when one spot on an array equals one
gene or one open reading frame, since sequence information is available.
This sequence represents a genomic sequence tag (GST) used in the method
of the invention. Note: The sequence data for this patent did not form
part of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1368 BP; 314 A; 365 C; 329 G; 360 T; 0 U; 0 Other;
SQ
Query Match 71.2%; Score 17.8; DB 6; Length 1368;
Best Local Similarity 90.5%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAACGACACAGCGAGCCCT 22
DB 1250 AAAAAACGACAAAGGGAGCCCT 1230
RESULT 6
ADG32053/C
ID ADG32053 standard; DNA; 2347 BP.
XX
XX AC ADG32053;
XX
XX DT 26-FEB-2004 (first entry)
XX
XX DE DNA encoding a mutant B_licheniformis secreted polypeptide SeqID 23.
XX
XX mutant; host cell; production yield; shelf life; product stability;
XX purity; secreted; gene; ds.
XX
XX OS Synthetic.
XX OS Bacillus licheniformis.
XX
XX PN WO2003093453-A2.
XX
XX PD 13-NOV-2003.
XX
XX PF 25-MAR-2003; 2003WO-DK000198.
XX
XX PR 10-APR-2002; 2002DK-00000534.
XX
XX (NOVO ) NOVOZYMES AS.
XX
XX Andersen JT, Jorgensen ST, Rasmussen MD, Olsen PB, Clausen IG;
XX
XX WPI; 2004-053045/05.
XX P-PSDB; ADG32054.
XX
XX New mutant Bacillus licheniformis host cell secreting 5 % less of one or
XX more secreted polypeptides than the parent host cell, useful for
XX producing a product of interest e.g. polypeptides, amino acids or
XX carbohydrates.
XX
XX Disclosure; SEQ ID NO 23; 422pp; English.
XX
XX This invention relates to a novel Bacillus licheniformis (B.
XX licheniformis) mutant host cell derived from a parent B. licheniformis
XX host cell that is mutated in genes encoding secreted polypeptides. host
XX Specifically, it refers to the generation of an improved Bacillus host
XX cell that reduces the need for product purification caused by contaminant
XX secreted native polypeptides in the culture medium. Accordingly, the
XX present invention describes reducing the expression of these native
```

CC proteins (e.g. proteolytic enzymes, nutrient uptake factors and signal  
 CC molecule), which in turn makes it easier to purify the heterologous  
 CC product of interest and therefore improving the production process.  
 CC Further benefits of a mutated host cell include an increase in total  
 CC production yield and a longevity of shelf life attributable to improved  
 CC product stability and purity. This polynucleotide is a DNA sequence  
 CC encoding a mutant B. licheniformis secreted polypeptide of the invention.

XX Sequence 2347 BP; 681 A; 567 C; 542 G; 557 T; 0 U; 0 Other;

Query Match 71.2%; Score 17.8; DB 12; Length 2347;  
 Best Local Similarity 90.5%; Pred. No. 2.8e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACGAGCCCT 22

Db 327 AAAAAGCAGACGAGCCCT 307

RESULT 7

ABD32610/C

ID ABD32610 standard; DNA; 115780 BP.

XX AC ABD32610;

XX DT 18-NOV-2004 (first entry)

XX DE Mouse cancer-associated genomic DNA MD12-017.

XX KW Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;

XX KW leukaemia; lymphoma; CAP.

XX OS Mus musculus.

XX PN WO2004074320-A2.

XX PD 02-SEP-2004.

XX PF 17-FEB-2004; 2004WO-US004730.

XX PR 14-FEB-2003; 2003US-00367094.

XX PR 14-MAR-2003; 2003US-00388838.

XX PR 15-APR-2003; 2003US-00417375.

XX PR 13-JUN-2003; 2003US-00461862.

XX PR 15-SEP-2003; 2003US-00663431.

XX PR 15-DEC-2003; 2003US-00737318.

XX PA (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Morris DW, Malandro MS;

XX DR WPI; 2004-652914/63.

XX PT New isolated cancer-associated polynucleotides and polypeptides useful

XX PT for diagnosing, preventing or treating cancers, especially lymphoma and

XX PT leukemia, or in screening for agents that modulate cancer.

XX PS disclosure; seqid 122; 310pp; English.

XX CC The invention relates to an isolated nucleic acid comprising at least 10

XX CC contiguous nucleotides of any of the 233 polynucleotide sequences given

XX CC in the specification, or its complement. The nucleic acids encode cancer-

XX CC associated proteins. Also included are an expression vector comprising

XX CC the isolated nucleic acid cited above, a host cell comprising the above

XX CC recombinant nucleic acid or expression vector, a microarray for detecting

XX CC a cancer-associated (CA) nucleic acid comprising at least one probe

XX CC comprising at least 10 contiguous nucleotides of any of the above-

XX CC mentioned nucleotide sequences, an isolated polypeptide (encoded within

XX CC an open reading frame of a CA sequence selected from any of the 95

XX CC polynucleotide sequences as mentioned in the specification, or its

XX CC complement), an isolated antibody, (or its antigen binding fragment) that

XX CC binds to the above polypeptide, a hybridoma that produces the above

XX CC monoclonal antibody, a pharmaceutical composition comprising the above

CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
 CC cells(comprising the antibody cited above, methods for diagnosing cancer  
 CC or for detecting the presence or absence of cancer cells in an  
 CC individual, a method for inhibiting growth of cancer cells in an  
 CC individual, a method for delivering a therapeutic agent to cancer cells  
 CC in an individual, an electronic library comprising the above  
 CC polynucleotide or polypeptide (or their fragments), methods of screening  
 CC for anticancer activity or for a bioactive agent capable of modulating  
 CC the activity of a CA protein (CAP), methods for detecting cancer  
 CC associated with expression of a polypeptide in a test cell sample, a  
 CC method for treating cancers and a method for inhibiting the expression of  
 CC CA gene in a cell. The composition and methods are useful for detecting,  
 CC diagnosing, preventing and treating cancers, especially lymphoma and  
 CC leukaemia. These may also be used in screening for agents that modulate  
 CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 115780 BP; 27988 A; 26066 C; 28294 G; 33243 T; 0 U; 189 Other;

Query Match 71.2%; Score 17.8; DB 13; Length 115780;

Best Local Similarity 90.5%; Pred. No. 4.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGCAGACAGCGAGCCCTCGG 25

Db 111059 AAGCAGACAGCGAGCCCTCGG 111039

RESULT 8

ABK79534

ID ABK79534 standard; DNA; 559 BP.

XX AC ABK79534;

XX DT 13-AUG-2002 (first entry)

XX DE Bacillus clausii genomic sequence tag (GST) #2377.

XX KW Differential gene expression; genomic sequenced tag; GST;

XX KW altered culture condition; environmental stress;

XX KW physiological provocation; ds.

XX OS Bacillus clausii.

XX PN WO200229113-A2.

XX PD 11-APR-2002.

XX PF 05-OCT-2001; 2001WO-US031437.

XX PR 06-OCT-2000; 2000US-00680598.

XX PR 27-MAR-2001; 2001US-0279526P.

XX PA (NOVO ) NOVOZYMES BIOTECH INC.

XX PA (NOVO ) NOVOZYMES AS.

XX PI Berka R, Clausen IG;

XX DR WPI; 2002-416684/44.

XX PT Monitoring differential expression of several genes in first Bacillus

XX PT cell relative to expression of same genes in one or more second Bacillus

XX PT cells, by using substrate containing Bacillus genomic sequenced tag

XX PS Claim 11; SEQ ID NO 6825; 200pp; English.

XX CC The invention describes a method of monitoring differential expression of

XX CC genes in a first Bacillus cell relative to expression of the genes in

XX CC other Bacillus cells, comprising hybridising labelled nucleic acid probes

XX CC isolated from Bacillus cells to a substrate containing array of Bacillus



CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions,  
 CC environmental stress or other physiological provocation. Extensive follow  
 CC up characterisation is unnecessary, when one spot on an array equals one  
 CC gene or one open reading frame, since sequence information is available.  
 CC This sequence represents a genomic sequence tag (GST) used in the method  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 559 BP; 128 A; 132 C; 142 G; 157 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 559;  
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCG 24  
 |||||  
 Db 203 AAAAAACGACACATCGAAGCTTCG 226

RESULT 9  
 ABK79471  
 ID ABK79471 standard; DNA; 597 BP.  
 XX  
 AC ABK79471;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Bacillus clausii genomic sequence tag (GST) #2314.  
 XX  
 KW Differential gene expression; genomic sequenced tag; GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.  
 XX  
 OS Bacillus clausii.  
 XX  
 PN WO200229113-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 05-OCT-2001; 2001WO-US031437.  
 XX  
 PR 06-OCT-2000; 2000US-00680598.  
 PR 27-MAR-2001; 2001US-0279526P.  
 XX  
 PA (NOVO ) NOVOZYMES BIOTECH INC.  
 PA (NOVO ) NOVOZYMES AS.  
 XX  
 PI Berka R, Clausen IG;  
 XX  
 XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second Bacillus  
 PT cells, by using substrate containing Bacillus genomic sequenced tag  
 PT array.  
 XX  
 XX Claim 11; SEQ ID NO 6762; 200pp; English.

XX The invention describes a method of monitoring differential expression of  
 XX genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining

CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions,  
 CC environmental stress or other physiological provocation. Extensive follow  
 CC up characterisation is unnecessary, when one spot on an array equals one  
 CC gene or one open reading frame, since sequence information is available.  
 CC This sequence represents a genomic sequence tag (GST) used in the method  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 597 BP; 141 A; 143 C; 152 G; 161 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 597;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCG 24  
 |||||  
 Db 243 AAAAAACGACACATCGAAGCTTCG 266

RESULT 10  
 ABQ39709  
 ID ABQ39709 standard; DNA; 753 BP.  
 XX  
 AC ABQ39709;  
 XX

DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26300.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for  
 PT diagnosis and prognosis, comprises selective hybridization of amplicons  
 PT from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one member,  
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC degree of hybridisation to both classes is determined from the label on  
 CC the amplicon. From the ratio of labels hybridised to the two classes of  
 CC oligomers, the degree of methylation is calculated. The method is used:  
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention

SQ Sequence 753 BP; 273 A; 303 C; 111 G; 66 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 6; Length 753;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTCG 24  
 |||||  
 Db 286 AAAAAACGACCGCGACCCCGCG 309

RESULT 11  
 ABQ39708/c  
 ID ABQ39708 standard; DNA; 753 BP.  
 XX  
 AC ABQ39708;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26299.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP010074.  
 XX  
 PR 01-SEP-2000; 2000DE-01043826.  
 PR 05-SEP-2000; 2000DE-01044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;  
 WPI; 2002-371829/40.  
 Determining the degree of cytosine methylation in genomic DNA, useful for  
 diagnosis and prognosis, comprises selective hybridization of amplicons  
 from chemically treated DNA.

Claim 12; 56pp + Sequence Listing; 56pp; German.  
 This invention describes a novel method for determining the degree of  
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 genomic sample of DNA. The sample is treated chemically to convert  
 cytosine (C) but not methylated C, to uracil, then part of the genomic  
 DNA that contains the target C is amplified to form a labeled amplicon.  
 The amplicon is hybridised to two classes, each with at least one member,  
 of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 degree of hybridisation to both classes is determined from the label on  
 the amplicon. From the ratio of labels hybridised to the two classes of  
 oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention

SQ Sequence 753 BP; 66 A; 111 C; 303 G; 273 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 6; Length 753;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTCG 24  
 |||||  
 Db 468 AAAAAACGACCGCGACCCCGCG 445

RESULT 12  
 ABQ41506/c  
 ID ABQ41506 standard; DNA; 1017 BP.  
 XX  
 AC ABQ41506;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28097.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP010074.  
 XX  
 PR 01-SEP-2000; 2000DE-01043826.  
 PR 05-SEP-2000; 2000DE-01044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;  
 WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for  
 diagnosis and prognosis, comprises selective hybridization of amplicons  
 from chemically treated DNA.

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of  
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 genomic sample of DNA. The sample is treated chemically to convert  
 cytosine (C) but not methylated C, to uracil, then part of the genomic  
 DNA that contains the target C is amplified to form a labeled amplicon.  
 The amplicon is hybridised to two classes, each with at least one member,  
 of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 degree of hybridisation to both classes is determined from the label on  
 the amplicon. From the ratio of labels hybridised to the two classes of  
 oligomers, the degree of methylation is calculated. The method is used:  
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention

XX Sequence 1017 BP; 83 A; 148 C; 350 G; 436 T; 0 U; 0 Other;  
SQ

Query Match 70.4%; Score 17.6; DB 6; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 134 AAAAAACGACACGAGTACCCACG 111

RESULT 13  
ABQ41507  
ID ABQ41507 standard; DNA; 1017 BP.  
XX  
AC ABQ41507;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28098.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug, side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
XN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP010074.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
PR 05-SEP-2000; 2000DE-01044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
DR

Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX

Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation

CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention

XX Sequence 1017 BP; 436 A; 350 C; 148 G; 83 T; 0 U; 0 Other;  
SQ

Query Match 70.4%; Score 17.6; DB 6; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 884 AAAAAACGACACGCGTACCCACG 907

RESULT 14  
ABV28495  
ID ABV28495 standard; cDNA; 6073 BP.  
XX  
XX AC ABV28495;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 28486.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US005171.  
XX  
PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
DR

Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX

Claim 1; Page 5951-5952; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX

Sequence 6073 BP; 1593 A; 1457 C; 1588 G; 1428 T; 0 U; 7 Other;  
SQ

Query Match 70.4%; Score 17.6; DB 5; Length 6073;  
Best Local Similarity 83.3%; Pred. No. 3.9e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTCG 24  
||||| ||| |||||  
Db 12 AAAAAAGAGACTCGCGAGCCCTCG 35

## RESULT 15

ABV22669

ID ABV22669 standard; cDNA; 6073 BP.

XX AC

XX ABV22669;

XX AC

DT 13-SEP-2002 (first entry)

XX XX

DE Human prostate expression marker cDNA 22660.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

KW Homo sapiens.

OS WO200160860-A2.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 03-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PA Schlegel R, Endege WO, Monahan JE;

PI WPI; 2001-662795/76.

DR Novel isolated nucleic acid molecule associated with cancerous state of

XX prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 3977-3978; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)

CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 6073 BP; 1593 A; 1457 C; 1588 G; 1428 T; 0 U; 7 Other;

QY Query Match 70.4%; Score 17.6; DB 5; Length 6073;

Best Local Similarity 83.3%; Pred. NO. 3.9e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTCG 24

||||| ||| |||||

Db 12 AAAAAAGAGACTCGCGAGCCCTCG 35

RESULT 16

ADQ97433

ID ADQ97433 standard; DNA; 295772 BP.

XX AC

XX ADQ97433;

XX 07-OCT-2004 (first entry)

DT Human cancer associated sequence HD08-047, SEQ ID 410.

XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.

KW Homo sapiens.

OS WO2004060304-A2.

XX WO2004060304-A2.

XX 22-JUL-2004.

XX 22-DEC-2003; 2003WO-US041389.

XX 27-DEC-2002; 2002US-00330773.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-543781/52.

XX New isolated cancer associated nucleic acids comprising at least 10

PT contiguous nucleotides, useful for diagnosing, preventing and/or treating

PT cancers such as leukemia and lymphoma.

XX Claim 1; SEQ ID NO 410; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-

CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or

CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence

CC data for this patent did not form part of the printed specification, but

CC was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 295772 BP; 74399 A; 61916 C; 68604 G; 90853 T; 0 U; 0 Other;

QY Query Match 70.4%; Score 17.6; DB 12; Length 295772;

Best Local Similarity 83.3%; Pred. No. 6.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGCCCTCG 24

||||| ||| |||||

Db 9743 AAAAAAGAGACTCGCGAGCCCTCG 9766

RESULT 17

ADT45037

ID ADT45037 standard; cDNA; 1797 BP.

XX AC

XX ADT45037;

DT 02-DEC-2004 (first entry)

XX Bacterial polynucleotide #19788.

XX Recombinant DNA construct; transformed plant; improved plant property;

XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

XX pathogen tolerance; pest tolerance; plant disease resistance;

XX cell cycle pathway modification; plant growth regulator;

XX homologous recombination; seed oil yield; protein yield; carbohydrate;

XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX bacterial polynucleotide; gene; ss.

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

```
XX 20-FEB-2003; 2003US-00369493.
PF PA
XX 21-FEB-2002; 2002US-0360039P.
PR PA
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 43475; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1797 BP; 514 A; 332 C; 408 G; 543 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 13; Length 1797;
Best Local Similarity 86.4%; Pred. No. 5.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACGACACAGCGGCCT 22
Db 336 AAAAAACGACACAGCGGCCT 357
XX
RESULT 18
ACL38755
ID ACL38755 standard; cDNA; 2000 BP.
XX
XX ACL38755;
XX
XX 02-JUN-2005 (first entry)
XX
XX Rice stress-regulated promoter SEQ ID NO:17318.
XX
XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX agriculture.
XX
XX Oryza sativa.
XX
XX WO2003008540-A2.
XX
XX 30-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019668.
XX
XX 22-JUN-2001; 2001US-0300112P.
XX 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA
XX
XX Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Rieke D, Zhu T;
PI WPI; 2003-248011/24.
XX
XX New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
XX Claim 48; SEQ ID NO 17318; 89pp; English.
XX
XX The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
XX Sequence 2000 BP; 558 A; 432 C; 413 G; 597 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 11; Length 2000;
Best Local Similarity 86.4%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAACGACACAGCGGCCT 22
Db 1636 AAAAAACGACACAGCGGCCT 1657
XX
RESULT 19
ADB84162/c
ID ADB84162 standard; DNA; 6240 BP.
XX
XX ADB84162;
XX
XX 04-DEC-2003 (first entry)
XX
XX
XX Gene expression estimation method-related DNA sequence #22.
XX
XX gene expression; plant; flower appearance; ds.
XX
XX Unidentified.
XX
XX WO2003044227-A1.
XX
XX 30-MAY-2003.
XX
XX 21-NOV-2001; 2001WO-JP010195.
XX
XX 21-NOV-2001; 2001WO-JP010195.
XX
XX (NAAAG-) NAT INST AGROBIOLOGICAL SCI.
XX
```

PI Higo K, Iwamoto M;  
 XX WPI; 2003-449821/42.  
 DR  
 XX  
 XX Identification of sequences homologous to a transposable element in genes  
 XX of a plant such as rice for screening genes expressed in particular  
 PT organs of the plant including the flower.  
 PT  
 XX  
 XX Disclosure; Page 127-131; 147pp; Japanese.  
 XX  
 CC The invention comprises a method of estimating gene expression in a  
 CC chosen organ of a plant. The method involves searching sequences in the  
 CC neighbourhood of putative protein-encoding regions of genes expressed in  
 CC the organ which are homologous to a known transposon key sequence, and  
 CC then selecting for genes in which these sequences are present. The method  
 CC of the invention is useful for establishing the pattern of gene  
 CC expression in particular organs of a plant (e.g. the flower) in  
 CC connection with the improvement of varieties of rice and other plants and  
 CC modification of the structural form and appearance of flowers. The  
 CC present DNA sequence was used in the exemplification of the invention.  
 XX  
 XX Sequence 6240 BP; 1574 A; 1348 C; 1317 G; 2001 T; 0 U; 0 Other;  
 SQ  
 Query Match 68.8%; Score 17.2; DB 10; Length 6240;  
 Best Local Similarity 86.4%; Pred. No. 6e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAACGACACGAGCCCT 22  
 Db 1786 AAAAAATGACACACGACCAT 1765  
 RESULT 20  
 AAF44516/c  
 ID AAF44516 standard; cDNA; 283 BP.  
 XX  
 AC AAF44516;  
 XX  
 XX 26-MAR-2001 (first entry)  
 DT  
 XX  
 XX Mouse dextran sodium sulphate induced colitis EST SEQ ID NO:46.  
 DE  
 XX  
 XX Mouse; gastrointestinal inflammation; DSS-induced colitis; EST;  
 KW dextran sodium sulphate induced colitis; expressed sequence tag;  
 KW inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
 KW immunosuppressive; immunostimulant; antiarthritic; antirheumatic;  
 KW antianaemic; antiasthmatic; antidepressant; cerebroprotective; cardiac;  
 KW antiallergic; antiinflammatory; antiviral; antibacterial; antifungal;  
 KW vulnary; haematopoietic cell stimulator; differentiation; anaemia;  
 KW proliferation; haematopoietic cell; immunological deficiency syndrome;  
 KW blood disorder; wound healing; stroke; heart attack; asthma;  
 KW autoimmune disorder; rheumatoid arthritis; allergic reaction;  
 KW graft versus host disease; infectious disease; infection; ss.  
 XX  
 XX Mus musculus.  
 OS  
 XX  
 XX WO200077166-A2.  
 PN  
 XX  
 XX 21-DEC-2000.  
 PD  
 XX  
 XX 09-JUN-2000; 2000WO-US015973.  
 PP  
 XX  
 XX 10-JUN-1999; 99US-0138487P.  
 PR  
 XX  
 XX (DIGI-) DIGITAL GENE TECHNOLOGIES INC.  
 PA  
 XX  
 XX Viney J, Sims JE, Dubose RF, Hilbush BS, Hasel KW, Buchner RP;  
 PI  
 XX  
 XX WPI; 2001-061869/07.  
 DR  
 XX  
 XX New isolated nucleic acid molecule useful for prevention, treatment or  
 PT amelioration of a medical condition, e.g. inflammatory bowel disease.  
 PT  
 XX

PS Claim 1; Page 111; 128pp; English.  
 XX  
 XX AAF44471 to AAF44532 represent expressed sequences tags (ESTs) isolated  
 CC from a mouse having dextran sodium sulphate (DSS)-induced colitis (I).  
 CC AAB51241 represents a protein (II) encoded by (I). (I) can have  
 CC immunosuppressive; immunostimulant; antiarthritic; antirheumatic.  
 CC antianaemic; cardiac; antiasthmatic; antidepressant; cerebroprotective;  
 CC antiallergic; antiinflammatory; antiviral; antibacterial; antifungal and  
 CC vulnary activities, and are haematopoietic cell stimulators. (I) and  
 CC (II) can be used to prevent, treat or ameliorate a medical condition. (I)  
 CC and (II) can be used to increase differentiation and proliferation of  
 CC haematopoietic cells in the treatment of immunological deficiency  
 CC syndromes, blood disorders (e.g. anaemia, wound healing, stroke, heart  
 CC attack), autoimmune disorders (e.g. rheumatoid arthritis, asthma,  
 CC allergic reactions, modulate inflammation, especially inflammatory bowel  
 CC disease and Crohn's disease), graft versus host disease and infectious  
 CC diseases (e.g. viral, bacterial or fungal infection). AAF44533 to  
 CC AAF44598 represent PCR primers used in the exemplification of the present  
 CC invention  
 XX  
 XX Sequence 283 BP; 62 A; 74 C; 66 G; 81 T; 0 U; 0 Other;  
 SQ  
 Query Match 68.0%; Score 17; DB 4; Length 283;  
 Best Local Similarity 80.0%; Pred. No. 5.2e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 AAAAAACGACACGAGCCCTCGG 25  
 Db 54 AAAAAACGGGTCTATCGGCGCTCGG 30  
 RESULT 21  
 ABD00388/c  
 ID ABD00388 standard; DNA; 603 BP.  
 XX  
 AC ABD00388;  
 XX  
 XX 29-JUL-2004 (first entry)  
 DT  
 XX  
 XX Klebsiella pneumoniae polynucleotide seqid 6163. .  
 DE  
 XX  
 XX Recombinant expression vector; transcription regulatory element;  
 KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.  
 KW  
 XX Klebsiella pneumoniae.  
 OS  
 XX US6610836-B1.  
 PN  
 XX  
 XX 26-AUG-2003.  
 PD  
 XX  
 XX 27-JAN-2000; 2000US-00489039.  
 PF  
 XX  
 XX 29-JAN-1999; 99US-0117747P.  
 PR  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX  
 XX Breton GL, Osborne M;  
 PI  
 XX  
 XX WPI; 2003-895346/82.  
 DR  
 XX P-PSDB; ABO66817.  
 DR  
 XX  
 XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
 PT preparing a vaccine composition against Klebsiella pneumoniae.  
 PT  
 XX  
 XX Disclosure; SEQ ID NO 6163; 932pp; English.  
 PS  
 XX  
 XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
 CC pneumoniae polypeptide. Also described are: a recombinant expression  
 CC vector comprising the nucleic acid, operably linked to a transcription  
 CC regulatory element; and a cell comprising the recombinant expression  
 CC vector. The nucleic acid is useful for preparing a vaccine composition  
 CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
 CC pneumoniae polypeptide of the invention

```
XX SQ Sequence 603 BP; 139 A; 182 C; 182 G; 100 T; 0 U; 0 Other;
Query Match 68.0%; Score 17; DB 11; Length 603;
Best Local Similarity 80.0%; Pred. No. 5.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25
    ||||| ||||| ||||| ||||| |||||
Db 164 AAATAGCGAAGACAGCGTTCCTCGG 140

RESULT 22
AA43897/C
ID AAA43897 standard; cDNA; 746 BP.
XX AC AAA43897;
XX DT 21-AUG-2000 (first entry)
XX DE Human secreted expressed sequence tag SEQ ID NO:472.
XX KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
XX antitumor; osteoprotective; neuroprotective; nontropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
XX autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
XX infection; depression; psoriasis; ss.
XX OS Homo sapiens.
XX PN WO200021991-A1.
XX PD 20-APR-2000.
XX PF 15-OCT-1999; 99WO-US024206.
XX PR 15-OCT-1998; 98US-0104436P.
XX PA (GEM ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Bowman MR;
XX WPI; 2000-317938/27.
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (sESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders.
XX Claim 1; Page 327-328; 803pp; English.
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
XX sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue
XX sources. The sESTs can have a range of activities depending on the
XX tissues they were isolated from. The activities include: chemotactic;
XX proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
XX haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
XX antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor;
XX osteoprotective; neuroprotective; nontropic; antiparkinsonian; antipsoriatic;
XX cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be
XX used for gene therapy and in vaccines. The sESTs are useful as probes for
XX the identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the sESTs. Proteins encoded by the sESTs
XX are useful in assays for determining biological activity and raising
```

```
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention
XX SQ Sequence 746 BP; 174 A; 190 C; 181 G; 199 T; 0 U; 2 Other;
Query Match 68.0%; Score 17; DB 3; Length 746;
Best Local Similarity 80.0%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25
    ||||| ||||| ||||| ||||| |||||
Db 446 AAAAAACCAACAAATGGAGCCCTCAG 422

RESULT 23
AAK84896
ID AAK84896 standard; DNA; 3271 BP.
XX AC AAK84896;
XX DT 07-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39708.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
```



```
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0231968P.
PR 13-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
17-NOV-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250319P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.

XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 39708; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 3271 BP; 959 A; 532 C; 679 G; 1101 T; 0 U; 0 Other;
XX
XX Query Match 68.0%; Score 17; DB 4; Length 3271;
XX Best Local Similarity 80.0%; Pred. NO. 6.9e+02;
XX Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25
Db 2473 AAAAAAAGACACAGCGAAGCACAG 2497
```

```
RESULT 24
ADR07098
ID ADR07098 standard; cDNA; 4473 BP.
XX AC
XX ADR07098;
AC ADR07098;
XX DT
XX 04-NOV-2004 (first entry)
XX DE
XX Full length human cDNA useful for treating neurological disease Seq 604.
XX KW
XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW tranquiliser.
XX KW
XX Homo sapiens.
XX OS
XX Homo sapiens.
XX PF
XX EP1447413-A2.
XX PN
XX 18-AUG-2004.
XX PD
XX 12-FEB-2004; 2004EP-00003145.
XX PF
XX 14-FEB-2003; 2003JP-00102207.
XX PR
XX 09-MAY-2003; 2003JP-00131452.
XX PS
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX PA
XX Isozai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
PI WPI; 2004-583265/57.
DR P-PSDB; ADR09054.
XX KW
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases.
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
PT
XX Claim 1; SEQ ID NO 604; 2686pp; English.
XX PS
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to
CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC cytostatic and tranquiliser activities. This polynucleotide is a full
CC length human cDNA sequence of the invention. NOTE: This sequence is not
CC given in the sequence listing of the specification but can be obtained on
CC CD-ROM from the European Patent Office, Vienna Sub-office.
XX CD-ROM from the European Patent Office, Vienna Sub-office.
XX SQ
XX Sequence 4473 BP; 989 A; 1238 C; 1138 G; 1108 T; 0 U; 0 Other;
Query Match 68.0%; Score 17; DB 13; Length 4473;
Best Local Similarity 80.0%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AAAAAACGACACGAGCCCTCGG 25
DB 3640 AAAAAACGACACGAGCCCTCGG 3664
RESULT 25
AAS65021
ID AAS65021 standard; cDNA; 6092 BP.
XX AC
XX AAS65021;
XX DT
XX 13-FEB-2002 (first entry)
XX DE
XX DNA encoding novel human diagnostic protein #825.
XX KW
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS
XX Homo sapiens.
XX PN
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US008631.
XX PR
XX 31-MAR-2000; 2000US-00540217.
XX PR
XX 23-AUG-2000; 2000US-00649167.
XX PA
XX (HYSE-) HYSEQ INC.
XX PI
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG00834.
XX KW
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS
XX Claim 1; SEQ ID NO 825; 103pp; English.
XX CC
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques involving (II). (II) is
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 6092 BP; 1677 A; 1318 C; 1389 G; 1708 T; 0 U; 0 Other;
Query Match 68.0%; Score 17; DB 5; Length 6092;
Best Local Similarity 80.0%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AAAAAACGACACGAGCCCTCGG 25
DB 2200 AAAAAAGGAGATAGCGACCCCTCAG 2224
RESULT 26
AEA35989
ID AEA35989 standard; DNA; 7318 BP.
XX
```

```
AC AEA35989;
XX
XX
DT 25-AUG-2005 (first entry)
XX
XX DE Long terminal repeat (LTR) retrotransposon associated DNA SEQ ID NO 1.
XX
XX ds; long terminal repeat; LTR; retrotransposon; transgenic.
XX
XX Unidentified.
XX
XX OS
XX
XX PN WO2005054463-A1.
XX
XX 16-JUN-2005.
XX
XX PF 19-NOV-2004; 2004WO-JP017307.
XX
XX 21-NOV-2003; 2003JP-00393175.
XX
XX (OSAI-) OSAKA IND PROMOTION ORG.
XX PA (AGEN ) NAT INST RADIOLOGICAL SCI.
XX
XX Takeda J, Horie K, Yusa K, Ishihara H;
XX
XX WPI; 2005-435392/44.
XX
XX Novel isolated nucleic acid construct comprising sequence that encodes
XX PT long terminal repeat retrotransposon, useful for modifying genome,
XX PT conforming dislocation capability of retrotransposon or producing
XX PT transgenic organism.
XX
XX Claim 8; SEQ ID NO 1; 129pp; Japanese.
XX
XX The invention relates to an isolated nucleic acid construct comprising a
XX CC sequence that encodes a long terminal repeat (LTR) retrotransposon. The
XX CC construct is useful for modifying a genome. The foreign gene is arranged
XX CC in the reverse transcription direction of retrotransposon. The foreign
XX CC gene encodes a factor that provides resistance to antibiotics, nutrition
XX CC replenishment factor, enzyme or fluorescent. The construct is useful for
XX CC conforming whether the retrotransposon has dislocation capability. The
XX CC step of detection involves obtaining sequence obtained by ligation
XX CC mediated PCR and comparing the genomic database with the obtained
XX CC sequence. The construct is useful for producing transgenic organism. The
XX CC construct is useful for rearranging the foreign gene or transducing a
XX CC foreign gene into a host. The construct enables efficient modification of
XX CC a genome. The present sequence represents a long terminal repeat (LTR)
XX CC retrotransposon associated DNA.
XX
XX SQ Sequence 7318 BP; 1990 A; 1649 C; 1714 G; 1965 T; 0 U; 0 Other;
XX
XX Query Match 68.0%; Score 17; DB 14; Length 7318;
XX Best Local Similarity 80.0%; Pred. No. 7.6e+02;
XX Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
XX ||||| ||||| ||||| ||||| |||||
XX Db 2289 AAAAAACGGTCCACGGGCGCTCGG 2313
XX
XX RESULT 27
XX AAK87049
XX ID AAK87049 standard; DNA; 16605 BP.
XX
XX AC AAK87049;
XX
XX 07-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41861.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
```

PN WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.  
XX 14-SEP-2000; 2000US-0232399P.  
XX 14-SEP-2000; 2000US-0232400P.  
XX 14-SEP-2000; 2000US-0232401P.  
XX 14-SEP-2000; 2000US-0233063P.  
XX 14-SEP-2000; 2000US-0233064P.  
XX 14-SEP-2000; 2000US-0233065P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234997P.  
XX 25-SEP-2000; 2000US-0234998P.  
XX 26-SEP-2000; 2000US-0235484P.  
XX 27-SEP-2000; 2000US-0235834P.  
XX 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249277P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
XX PA  
XX PI

XX WPI; 2001-483426/52.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX Disclosure; SEQ ID NO 41861; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to prevent the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK97694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX SQ Sequence 16605 BP; 4009 A; 3696 C; 4212 G; 4688 T; 0 U; 0 Other;  
Query Match 68.0%; Score 17; DB 4; Length 16605;  
Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGCGAGCCCTCGG 25  
Db 1358 AAAAAAACACAGCCGCCAAGG 1382  
RESULT 28  
AAK65489  
ID AAK65489 standard; DNA; 16605 BP.  
XX AC AAK65489;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20301.  
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX OS cytostatic; gene therapy; vaccine; metastasis; ds.  
XX OS Homo sapiens.  
XX FN WO200157182-A2.  
XX PD 09-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001354.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.



CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK5950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 16605 BP; 4009 A; 3696 C; 4212 G; 4688 T; 0 U; 0 Other;

Query Match 68.0%; Score 17; DB 4; Length 16605;  
Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAGCAGCAGCAGCAGCCTCGG 25  
IIIIII IIIIIII IIIIIII IIIIIII  
Db 1358 AAAAAGCAGCAGCAGCAGCCTCGG 1382

RESULT 29  
AAK90339  
ID AAK90339 standard; DNA; 17431 BP.  
XX  
AC AAK90339;  
XX  
DT  
DT  
DE Human digestive system antigen genomic sequence SEQ ID NO: 3915.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO20015314-A2.  
XX

PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001324.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239353P.  
PR 13-OCT-2000; 2000US-023937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241828P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.

```
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
XX Disclosure; SEQ ID NO 3915; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
XX Sequence 17431 BP; 5614 A; 3722 C; 3532 G; 4563 T; 0 U; 0 Other;
SQ
Query Match 68.0%; Score 17; DB 4; Length 17431;
Best Local Similarity 80.0%; Pred. No. 8.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AAAAAACGACACAGCGAGCCTCGG 25
Db 7370 AAAAAACAAAAATCGAGCCTCTG 7394
RESULT 30
AA157710
ID AA157710 standard; DNA; 17431 BP.
XX
AC AA157710;
XX
XX 19-OCT-2001 (first entry)
XX
XX Human colorectal cancer antigen coding sequence SEQ ID NO: 247.
DE Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
XX Homo sapiens.
OS
```

```
XX
PN WO200155350-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001350.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
```



PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457727/49.  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon and rectum including colorectal cancers  
PT and also for testing and detection e.g. diagnosis.  
XX Disclosure; SEQ ID NO 247; 522pp + Sequence Listing; English.  
XX The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in AA157547-  
CC AA757619 and AA38569-AA38641. These can be used in the diagnosis,  
CC prevention and treatment of cancer of the colon and/or rectum. The  
CC present sequence is a colorectal cancer antigen genomic sequence. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 17431 BP; 5614 A; 3722 C; 3532 G; 4563 T; 0 U; 0 Other;  
Query Match 68.0%; Score 17; DB 4; Length 17431;  
Best Local Similarity 80.0%; Pred. NO. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGCGAGCCTCGG 25  
Db 7370 AAAAAACAAATAATGCGAGCCTCTG 7394  
RESULT 31  
ABS99887  
ID ABS99887 standard; DNA; 17431 BP.  
XX  
AC ABS99887;  
XX 18-DEC-2002 (first entry)  
DT  
XX Genomic DNA #91 encoding human colorectal cancer related protein.  
DE  
XX Human; colorectal cancer related protein; colon; rectum;  
KW colorectal cancer metastasis; gastrointestinal disorder; cytostatic;  
KW gene; ds.  
XX Homo sapiens.  
OS  
XX US2002119919-A1.  
PN  
XX 29-AUG-2002.  
PD  
XX 17-JAN-2001; 2001US-00764855.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR  
XX (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2002-731367/79.  
DR  
XX New colorectal cancer polypeptide for diagnosing, prognosing, preventing,  
PT and treating immune, hyperproliferative, liver, kidney, reproductive  
PT disorders and for identifying modulators of therapeutic use.  
XX Disclosure; SEQ ID NO 247; 183pp; English.  
PS  
XX The present invention relates to the isolation of novel human colorectal  
CC cancer related proteins, and polynucleotide sequences encoding them. The  
CC sequences of the invention are useful in the diagnosis, treatment,  
CC prevention and/or prognosis of the colon and/or rectum, including  
CC colorectal cancer, colorectal cancer metastases, and gastrointestinal

CC disorders such as dysphagia, peptic oesophagitis, gastric reflux,  
 CC irritable bowel syndrome, and peritoneal diseases. The invention also  
 CC describes antibodies that bind colorectal cancer related proteins,  
 CC vectors, host cells, and recombinant and synthetic methods for producing  
 CC human colorectal cancer related polynucleotides, polypeptides, and/or  
 CC antibodies. ABS99797-ABS99974 represent genomic sequences encoding human  
 CC colorectal cancer related proteins. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/psipdDentry.html  
 XX

SQ Sequence 17431 BP; 5614 A; 3722 C; 3532 G; 4563 T; 0 U; 0 Other;

Query Match 68.0%; Score 17; DB 6; Length 17431;  
 Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
 ||||| | | ||||| |  
 Db 7370 AAAAAACAAAATGCGAGCCCTCTG 7394

RESULT 32

ADB93040

ID ADB93040 standard; DNA; 17431 BP.

XX

AC ADB93040;

XX

DT 04-DEC-2003 (first entry)

XX

DE Human colorectal cancer related polypeptide DNA #91.

XX

KW ds; gene; human; colorectal cancer; antigen; gene therapy;  
 KW gastrointestinal disorder; inflammatory disease; infection; cancer;  
 KW intestinal neoplasm; small intestine carcinoma; tumor;  
 KW small intestine non-Hodgkin's lymphoma; small bowel lymphoma; ulcer;  
 KW peptic ulcer; Bruton's disease; X linked infantile agammaglobulinemia;  
 KW severe combined immunodeficiency; DiGeorge anomaly;  
 KW hyperproliferative disorder; acute lymphoblastic leukaemia;  
 KW acute lymphocytic leukaemia; urinary system disorder; cortical necrosis;  
 KW kidney infarction; cardiovascular disorder; carcinoma heart disease;  
 KW arrhythmia; respiratory disorder; non-allergic rhinitis; sinusitis;  
 KW musculoskeletal system disorder; Albers-Schonberg disease;  
 KW Marfan's syndrome; neurological disease; phenylketonuria;  
 KW Wernicke's encephalopathy; Alzheimer's disease; endocrine disorder;  
 KW Grave's disease; Cushing's syndrome; reproductive system disorder;  
 KW prostatitis; benign prostatic hypertrophy; benign prostatic hyperplasia;  
 KW thrombosis; atherosclerosis; myocardial infarction; ischaemic attack.

OS Homo sapiens.

XX

PN US2003054420-A1.

XX

PD 20-MAR-2003.

XX

PF 11-FEB-2002; 2002US-00072349.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR

PR 04-FEB-2000; 2000US-0180628P.

PR

PR 24-FEB-2000; 2000US-0184664P.

PR

PR 02-MAR-2000; 2000US-0186350P.

PR

PR 16-MAR-2000; 2000US-0189874P.

PR

PR 17-MAR-2000; 2000US-0190076P.

PR

PR 18-APR-2000; 2000US-0198123P.

PR

PR 19-MAY-2000; 2000US-0205515P.

PR

PR 07-JUN-2000; 2000US-0209467P.

PR

PR 28-JUN-2000; 2000US-0214886P.

PR

PR 30-JUN-2000; 2000US-0215135P.

PR

PR 07-JUL-2000; 2000US-0216647P.

PR

PR 07-JUL-2000; 2000US-0216880P.

PR

PR 11-JUL-2000; 2000US-0217487P.

PR

PR 11-JUL-2000; 2000US-0217496P.

PR

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239315P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 PR 17-JAN-2001; 2001US-00764855.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2003-708345/67.  
 DR  
 XX Novel colorectal cancer antigen useful for treating, preventing,  
 PT diagnosing and/or prognosing gastrointestinal disorders, infections,  
 PT cancers such as intestinal neoplasms, ulcers.  
 XX  
 XX Disclosure; SEQ ID NO 247; 179pp; English.  
 PS  
 XX The invention relates to a colorectal cancer antigen. The antigen is  
 CC useful for chromosome identification, chromosome mapping, radiation  
 CC hybrid mapping or gene therapy, or as hybridisation probes for  
 CC differential identification of the tissues or cell types present in a  
 CC biological sample. The antigen is useful for treating, preventing,  
 CC diagnosing and/or prognosing gastrointestinal disorders, including  
 CC inflammatory diseases and/or conditions, infections, cancers (e.g.  
 CC intestinal neoplasms (carcinoid tumour of the small intestine, non-  
 CC Hodgkin's lymphoma of the small intestine, small bowel lymphoma)) and  
 CC ulcers (e.g. peptic ulcers). The antigen and its nucleic acid are useful  
 CC to provide immunological probes for differential identification of the  
 CC tissue. The antigen and its nucleic acid are useful for treating,  
 CC preventing, diagnosing and/or prognosing diseases, disorders and/or

CC conditions of the immune system e.g. Bruton's disease, X linked infantile  
 CC agammaglobulinaemia, severe combined immunodeficiencies, DiGeorge  
 CC anomaly, etc. The antigen and its nucleic acid is useful for treating,  
 CC preventing and/or diagnosing hyperproliferative disorders (e.g. acute  
 CC lymphoblastic leukaemia, acute lymphocytic leukaemia, etc), urinary

Query Match 68.0%; Score 17; DB 10; Length 17431;  
 Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 7370 AAAAAACAAAAATGCGAGCCCTCTG 7394

RESULT 33  
 ADQ97885/c  
 ID ADQ97885 standard; DNA; 62130 BP.  
 XX  
 AC ADQ97885;  
 XX  
 XX 07-OCT-2004 (first entry)  
 DT  
 DE Mouse cancer associated sequence MD11-019, SEQ ID 862.  
 XX  
 XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.  
 KW  
 XX Mus musculus.  
 OS  
 XX WO2004060304-A2.  
 PN  
 XX 22-JUL-2004.  
 PD  
 XX 22-DEC-2003; 2003WO-US041389.  
 PF  
 XX 27-DEC-2002; 2002US-00330773.  
 PR  
 XX (SAGR-) SAGRES DISCOVERY INC.  
 PA  
 XX Morris DW, Malandro MS;  
 PI  
 XX WPI; 2004-543781/52.  
 DR  
 XX New isolated cancer associated nucleic acids comprising at least 10  
 PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
 PT cancers such as leukemia and lymphoma.  
 XX  
 XX Claim 1; SEQ ID NO 862; 199pp; English.  
 PS  
 XX The present invention relates to cancer associated sequences (ADQ97025-  
 CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
 CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 62130 BP; 15683 A; 13686 C; 14510 G; 17994 T; 0 U; 257 Other;  
 SQ

Query Match 68.0%; Score 17; DB 12; Length 62130;  
 Best Local Similarity 80.0%; Pred. No. 9.8e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 18465 AGAAAAGGACACAGAGGGCTCTCGG 18441

RESULT 34  
 ABD33618  
 ID ABD33618 standard; DNA; 62278 BP.  
 XX  
 AC ABD33618;  
 XX

DT 18-NOV-2004 (first entry)  
XX Murine cancer-associated (CA) gene MD07-126.  
DE Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
KW cancer; cytostatic.  
XX Mus musculus.  
OS WO2004058146-A2.  
PN 15-JUL-2004.  
XX 15-DEC-2003; 2003WO-US040081.  
XX 17-DEC-2002; 2002US-00322281.  
PR (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Malandro MS;  
PI WPI; 2004-499109/47.  
XX Novel human cancer associated protein encoded within open reading frame  
PT of cancer associated gene, useful as targets for diagnosing cancer.  
XX Disclosure; SEQ ID NO 849; 182pp; English.  
PS The invention relates to cancer-associated proteins (CAP) and the cancer-  
CC associated (CA) nucleic acids encoding them. The invention also relates  
CC to a method for treating cancers involving administering to a patient an  
CC inhibitor of CAP, and a method of screening for anticancer activity in a  
CC potential drug involving providing a cell that expresses a CA gene,  
CC contacting a tissue sample derived from a cancer cell with an anticancer  
CC drug candidate and monitoring the effect of the anticancer drug candidate  
CC on expression of the CA gene. The CAP proteins are useful for detecting  
CC cancer associated with expression of a CAP protein in a test cell sample  
CC and for screening for a bioactive agent capable of modulating the  
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
CC cancer, involving determining the expression of a CA nucleic acid in a  
CC tissue. This sequence represents a murine CA gene of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 62278 BP; 19201 A; 12593 C; 12586 G; 16957 T; 0 U; 941 Other;  
Query Match 68.0%; Score 17; DB 13; Length 62278;  
Best Local Similarity 80.0%; Pred. No. 9.8e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 28443 AAAAAACGGTCATCGGGCCCTCGG 28467  
RESULT 35  
ADD14752/c  
ID ADD14752 standard; cDNA; 86765 BP.  
XX ADD14752;  
AC 01-JAN-2004 (first entry)  
XX Human src biomarker polynucleotide SEQ ID NO:146.  
DE predictor set; protein tyrosine kinase activity modulator;  
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;  
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.  
XX Homo sapiens.  
OS WO2003062395-A2.  
PN

XX 31-JUL-2003.  
PD 17-JAN-2003; 2003WO-US001981.  
PF 18-JAN-2002; 2002US-0350061P.  
PR (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Huang F, Fairchild CR, Lee FY, Shaw P;  
PI WPI; 2003-636735/60.  
XX P-PSDB; ADD14150.  
DR New polynucleotides and polypeptides for predicting the activity of  
XX compounds that interact with protein tyrosine kinases and/or protein  
XX tyrosine kinase pathways.  
XX Claim 2; SEQ ID NO 146; 139pp; English.  
XX The present invention describes a predictor set comprising a plurality of  
XX polynucleotides or polypeptides whose expression pattern is predictive of  
XX the response of cells to treatment with a compound that modulates protein  
XX tyrosine kinase activity or members of the protein tyrosine kinase  
XX pathway. Also described: (1) predicting whether a compound is capable of  
XX modulating the activity of cells, comprising obtaining a sample of cells,  
XX determining whether the cells express a plurality of markers, and  
XX correlating the expression of the markers to the compound's ability to  
XX modulate the activity of the cells; (2) a plurality of cell lines for  
XX identifying polynucleotides and polypeptides whose expression levels  
XX correlate with compound sensitivity or resistance of cells associated  
XX with a disease state; and (3) identifying polynucleotides and  
XX polypeptides that predict compound sensitivity or resistance of cells  
XX associated with a disease state, comprising subjecting the plurality of  
XX cell lines to one or more compounds, analysing the expression pattern of  
XX a microarray of polynucleotides or polypeptides, and selecting  
XX resistance of cells associated with a disease state by using the  
XX expression pattern of the microarray. The polynucleotides and  
XX polypeptides have cytostatic activities, and can be used in gene therapy.  
XX The polynucleotides and polypeptides are useful in predicting the  
XX activity of compounds that interact with protein tyrosine kinases and/or  
XX protein tyrosine kinase pathways. These may be used in determining drug  
XX sensitivity in patients to allow the development of individualized  
XX genetic profiles which aid in treating diseases and disorders (e.g.  
XX cancer) based on patient response at a molecular level. The present  
XX sequence is used in the exemplification of the present invention.  
XX  
SQ Sequence 86765 BP; 24568 A; 21025 C; 20504 G; 20568 T; 0 U; 0 Other;  
Query Match 68.0%; Score 17; DB 10; Length 86765;  
Best Local Similarity 80.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 51218 AAAAAAACACAGCGAGCCCAAGG 51194  
RESULT 36  
ACN44296  
ID ACN44296 standard; DNA; 115284 BP.  
XX ACN44296;  
AC 18-NOV-2004 (first entry)  
XX Mouse genomic sequence MCG17138.  
DE Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX Mus musculus.  
OS



CC binds to a polypeptide, which is prepared by immunizing a host animal  
 CC with a composition comprising the polypeptide or its antigen binding  
 CC fragment and collecting cells from the host expressing antibodies against  
 CC the antigen or its antigen binding fragment, a composition comprising the  
 CC antibody and a carrier, a method of screening for anticancer activity, a  
 CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
 CC method of treating cancer and a method of inhibiting expression of a CA  
 CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
 CC nucleic acids. The antibody is useful for detecting the presence or  
 CC absence of cancer cells in an individual which involves contacting cells  
 CC from the individual with the antibody and detecting a complex of a CA  
 CC protein from the cancer cells and the antibody, where the detection of  
 CC the complex correlates with the presence of cancer cells in the  
 CC individual. The composition is useful for inhibiting growth of cancer  
 CC cells in an individual or for delivering a therapeutic agent to cancer  
 CC cells in an individual. The invention is also useful for diagnosing  
 CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
 CC a cell. This sequence represents murine cancer-associated genomic DNA of  
 CC the invention.

XX Sequence 163701 BP; 42438 A; 35611 C; 35831 G; 45904 T; 0 U; 3917 Other;  
 SQ

Query Match 68.0%; Score 17; DB 14; Length 163701;  
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCAGCAGCGCCCTCGG 25

Db 113080 AAAAAAGCAGCAGCGCCCTCGG 113056

RESULT 39

ABK83564/c

ID ABK83564 standard; cDNA; 198161 BP.

XX AC ABK83564;

XX 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #135.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US030821.

XX 03-OCT-2000; 2000US-0237189P.

XX (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.

XX Claim 1; SEQ ID NO 135; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation

CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where  
 CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) GA by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease using the gene expression  
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease, by detecting the level of  
 CC expression in a sample of the tissue of gene(s) from Gs, where the level  
 CC of expression of the gene is indicative of inflammation; (4) treating  
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease, by contacting a tissue having inflammation with an  
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
 CC is useful for detecting GCA; M2 is useful for modulating Gs; M3 is useful  
 CC for screening an agent capable of modulating GCA preferably in an  
 CC inflammation in a tissue; M4 is useful for detecting an inflammation  
 CC (especially chronic) in a tissue, an allergic response in a subject,  
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
 CC disease, ulcerative colitis, periodontal disease; also bacterial  
 CC infection, viral infection, parasitic infection, protozoal infection,  
 CC fungal infection and M5 is useful for treating one of the above  
 CC conditions. The present sequence represents a gene differentially  
 CC expressed in granulocytes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 198161 BP; 61561 A; 45946 C; 41360 G; 49294 T; 0 U; 0 Other;

Query Match 68.0%; Score 17; DB 6; Length 198161;

Best Local Similarity 80.0%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCAGCAGCGCCCTCGG 25

Db 78717 AAAAAAGCAGCGCTTGCCTAGG 78693

RESULT 40

ADQ17348/c

ID ADQ17348 standard; DNA; 198161 BP.

XX AC ADQ17348;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 165.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW ds.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX

DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 165; 210pp; English.  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 198161 BP; 61561 A; 45946 C; 41360 G; 49294 T; 0 U; 0 Other;  
Query Match 68.0%; Score 17; DB 12; Length 198161;  
Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCCTCGG 25  
Db 78717 AAAAAAGACCAGCTTGCCCTAGG 78693  
RESULT 41  
ADR52701/c  
ID ADR52701 standard; DNA; 198161 BP.  
XX  
AC ADR52701;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Drug therapy altered expressed gene #52.  
XX  
KW drug activity monitoring; expression profile; gene expression;  
KW peripheral blood sample; peripheral blood mononuclear cell; drug therapy;  
KW CCI-779; immunosuppressant; rapamycin; mammalian target of rapamycin;  
KW mTOR; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO2004072265-A2.  
XX  
PD 26-AUG-2004.  
XX  
PF 11-FEB-2004; 2004WO-US004118.  
XX  
PR 11-FEB-2003; 2003US-0446133P.  
PR 03-APR-2003; 2003US-0459782P.  
PR 23-JAN-2004; 2004US-0538246P.  
XX  
PA (AMHP ) WYETH.  
PA (BURC/) BURCZYNSKI M.  
PA (TWIN/) TWINE N.  
PA (DORN/) DORNER A J.  
PA (TREP/) TREPICCHIO W L.  
XX  
PI Burczynski M, Twine N, Dornier AJ, Trepicchio WL;  
XX  
DR WPI; 2004-642301/62.  
XX  
PT Monitoring drug activities in vivo comprises comparing an expression  
PT profile of a gene in a peripheral blood sample of a patient before and

PT after drug therapy.  
XX  
PS Disclosure; SEQ ID NO 52; 136pp; English.  
XX  
CC The invention relates to a method of monitoring drug activities in vivo  
CC by comparing an expression profile of at least one gene in a peripheral  
CC blood sample of a patient to a reference expression profile of the at  
CC least one gene, where the at least one gene is differentially expressed  
CC in peripheral blood mononuclear cells (PBMCs) of patients who have a non-  
CC blood disease and are subjected to a drug therapy as compared to PBMCs  
CC isolated from the patient before the drug therapy, and where the patient  
CC has the non-blood disease and is being treated by the drug therapy. The  
CC method, kit, and nucleic acid array are useful for monitoring drug  
CC activities in vivo. The drug is especially CCI-779, an ester analogue of  
CC the immunosuppressant rapamycin which is a potent inhibitor of the  
CC mammalian target of rapamycin (mTOR). This sequence represents a gene  
CC expressed in PBMC altered by the drug therapy. (Note: this sequence does  
CC no form part of the printed specification but was obtained in electronic  
CC format from WIPO at ftp.wipo.int/pub/published\_pct\_sequences/).  
XX  
SQ Sequence 198161 BP; 61561 A; 45946 C; 41360 G; 49294 T; 0 U; 0 Other;  
Query Match 68.0%; Score 17; DB 13; Length 198161;  
Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCCTCGG 25  
Db 78717 AAAAAAGACCAGCTTGCCCTAGG 78693  
RESULT 42  
ADR97430  
ID ADR97430 standard; DNA; 208765 BP.  
XX  
AC ADR97430;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Mouse cancer associated sequence MD08-047, SEQ ID 407.  
XX  
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.  
XX  
OS Mus musculus.  
XX  
FN WO2004060304-A2.  
XX  
PD 22-JUL-2004.  
XX  
PF 22-DEC-2003; 2003WO-US041389.  
XX  
PR 27-DEC-2002; 2002US-00330773.  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
PI Morris DW, Malandro MS;  
XX  
DR WPI; 2004-543781/52.  
XX  
PT New isolated cancer associated nucleic acids comprising at least 10  
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
PT cancers such as leukemia and lymphoma.  
XX  
PS Claim 1; SEQ ID NO 407; 199pp; English.  
XX  
CC The present invention relates to cancer associated sequences (ADR97025-  
CC ADR98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 208765 BP; 53491 A; 42141 C; 46142 G; 66771 T; 0 U; 220 Other;



Query Match 68.0%; Score 17; DB 12; Length 208765;  
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 10469 AAAAAAGAGACTCGGCGAGCCCTCGG 10493

RESULT 43  
 ADQ97362  
 ID ADQ97362 standard; DNA; 215126 BP.  
 XX  
 AC ADQ97362;  
 XX  
 DT 07-OCT-2004 (first entry)  
 XX  
 DE Mouse cancer associated sequence MD08-038, SEQ ID 339.  
 XX  
 KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO2004060304-A2.  
 XX  
 PD 22-JUL-2004.  
 XX  
 PF 22-DEC-2003; 2003WO-US041389.  
 XX  
 PR 27-DEC-2002; 2002US-00330773.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY INC.  
 XX  
 PI Morris DW, Malandro MS;  
 XX  
 DR WPI; 2004-543781/52.  
 XX  
 PT New isolated cancer associated nucleic acids comprising at least 10  
 FT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
 PT cancers such as leukemia and lymphoma.  
 XX  
 PS Claim 1; SEQ ID NO 339; 199pp; English.  
 XX  
 CC The present invention relates to cancer associated sequences (ADQ97025-  
 CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
 CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 215126 BP; 55133 A; 47649 C; 49476 G; 61330 T; 0 U; 1538 Other;

Query Match 68.0%; Score 17; DB 12; Length 215126;  
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 143743 AAGTAACGAAACACGAGCGGCTAGG 143767

RESULT 44  
 ADQ23222  
 ID ADQ23222 standard; DNA; 2217 BP.  
 XX  
 AC ADQ23222;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 6042.  
 XX  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

ds.  
 KW Homo sapiens.  
 OS  
 XX WO2004048938-A2.  
 PN  
 XX 10-JUN-2004.  
 PD  
 XX 26-NOV-2003; 2003WO-US038193.  
 PF  
 XX 26-NOV-2002; 2002US-0429739P.  
 PR  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 PA  
 XX Aziz N, Ginsburg WM, Zlotnik A;  
 PI  
 XX WPI; 2004-441208/41.  
 DR  
 XX Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX  
 XX Example 2; SEQ ID NO 6042; 210pp; English.  
 XX  
 CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX  
 SQ Sequence 2217 BP; 504 A; 592 C; 699 G; 422 T; 0 U; 0 Other;

Query Match 67.2%; Score 16.8; DB 12; Length 2217;  
 Best Local Similarity 90.0%; Pred. No. 8.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACGACACAGCGAGCCCTC 23  
 ||||| ||||| ||||| ||||| |||||  
 Db 163 AAACGACACAGCGAGCCCTC 182

RESULT 45  
 ABV14306  
 ID ABV14306 standard; cDNA; 354 BP.  
 XX  
 AC ABV14306;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 14297.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 XX  
 PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX WPI; 2001-662795/76.  
 XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.  
 PS Claim 1; Page 2387; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX Sequence 354 BP; 99 A; 92 C; 78 G; 85 T; 0 U; 0 Other;  
 SQ  
 Query Match 66.4%; Score 16.6; DB 5; Length 354;  
 Best Local Similarity 82.6%; Pred. No. 8.2e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAGACACAGCGAGCTCC 23  
 Db 44 AAAAAGACACAGCGAGCTCC 66  
 RESULT 46  
 ABV05137  
 ID ABV05137 standard; cDNA; 399 BP.  
 AC ABV05137;  
 XX 13-SEP-2002 (first entry)  
 DE Human prostate expression marker cDNA 5128.  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX Homo sapiens.  
 OS WO200160860-A2.  
 PN 23-AUG-2001.  
 PD 20-FEB-2001; 2001WO-US005171.  
 XX 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX

DR WPI; 2001-662795/76.  
 XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.  
 PS Claim 1; Page 871; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX Sequence 399 BP; 103 A; 107 C; 96 G; 92 T; 0 U; 1 Other;  
 SQ  
 Query Match 66.4%; Score 16.6; DB 5; Length 399;  
 Best Local Similarity 82.6%; Pred. No. 8.3e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAGACACAGCGAGCTCC 23  
 Db 79 AAAAAGACACAGCGAGCTCC 101  
 RESULT 47  
 ADQ20479  
 ID ADQ20479 standard; DNA; 401 BP.  
 XX ADQ20479;  
 XX 26-AUG-2004 (first entry)  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 3299.  
 XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW ds.  
 XX Homo sapiens.  
 OS WO2004048938-A2.  
 PN 10-JUN-2004.  
 PD 26-NOV-2003; 2003WO-US038193.  
 PF 26-NOV-2002; 2002US-0429739P.  
 PR (PROT-) PROTEIN DESIGN LABS INC.  
 PA Aziz N, Ginsburg WM, Zlotnik A;  
 PI WPI; 2004-441208/41.  
 DR Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.  
 XX Example 2; SEQ ID NO 3299; 210pp; English.  
 XX The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 401 BP; 98 A; 79 C; 95 G; 111 T; 0 U; 18 Other;  
SQ Sequence 401 BP; 98 A; 79 C; 95 G; 111 T; 0 U; 18 Other;  
Query Match 66.4%; Score 16.6; DB 12; Length 401;  
Best Local Similarity 79.2%; Pred. No. 8.3e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACAGCGCCCTCG 24  
Db 102 AAAAAACTACAAAGGANCCTCG 125

RESULT 48  
ACF85290  
ID ACF85290 standard; DNA; 412 BP.  
XX AC ACF85290;  
XX DT 02-JUN-2005 (first entry)  
XX DE Human SIRS/sepsis diagnostic marker DNA fragment 4150.  
XX KW Systemic inflammatory response syndrome; SIRS; antibacterial;  
XX KW immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.  
XX OS Homo sapiens.  
XX PN WO2004087949-A2.  
XX PD 14-OCT-2004.  
XX PF 31-MAR-2004; 2004WO-EP003419.  
XX PR 02-APR-2003; 2003DE-01015031.  
XX PR 08-AUG-2003; 2003DE-01036511.  
XX PR 02-SEP-2003; 2003DE-01040395.  
XX PA (SIRS-) SIRS LAB GMBH.  
XX PI Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;  
XX WPI; 2004-748070/73.  
XX DR In vitro detection of systemic inflammatory response syndrome and related  
XX PT conditions, for e.g. monitoring progression, comprises detecting abnormal  
XX PT expression of disease-related genes.  
XX PS Disclosure; Page; 75pp; German.

XX CC The invention relates to a novel method for in vitro detection of  
XX CC systemic inflammatory response syndrome (SIRS). The method comprises  
XX CC detecting abnormal expression of disease-related genes, or their  
XX CC associated peptides. The method of the invention demonstrates  
XX CC antibacterial, immunosuppressive and antiinflammatory applications and  
XX CC may be used for early differential diagnosis, monitoring progression, and  
XX CC assessing risk, assessing the likely response to treatment and for post  
XX CC mortem diagnosis of systemic inflammatory response syndrome, sepsis and  
XX CC sepsis-like conditions. The recombinant or synthetic nucleic acid  
XX CC sequences of the invention, or derived proteins or peptides, may be  
XX CC useful as calibrants in assays for the specified diseases, for evaluating  
XX CC activity or toxicity in screening for active agents and/or for  
XX CC preparation of agents for treatment or prevention of the specified  
XX CC diseases. The current sequence is that of a human SIRS/sepsis diagnostic  
XX CC marker DNA fragment of the invention. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at ftp.wipo.int/pub/published  
CC pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are  
CC disclosed within the specification, however, these have not been taken  
CC into account during indexing due to inconsistencies in application and  
CC format  
XX SQ Sequence 412 BP; 138 A; 61 C; 65 G; 148 T; 0 U; 0 Other;  
SQ Sequence 412 BP; 138 A; 61 C; 65 G; 148 T; 0 U; 0 Other;  
Query Match 66.4%; Score 16.6; DB 13; Length 412;  
Best Local Similarity 82.6%; Pred. No. 8.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACAGCGCCCTC 23  
Db 192 AAAAAATGATAGAGGAGCCCTC 214

RESULT 49  
ABV44224  
ID ABV44224 standard; cDNA; 459 BP.  
XX AC ABV44224;  
XX DT 16-SEP-2002 (first entry)  
XX DE Human prostate expression marker cDNA 44215.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX PD 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US005171.  
XX PR 17-FEB-2000; 2000US-0183319P.  
XX PR 16-MAR-2000; 2000US-0189862P.  
XX PR 25-MAY-2000; 2000US-0207454P.  
XX PR 09-JUN-2000; 2000US-0211314P.  
XX PR 18-JUL-2000; 2000US-0219007P.  
XX PR 13-DEC-2000; 2000US-0255281P.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX DR Novel isolated nucleic acid molecule associated with cancerous state of  
XX PT prostate cells and correlating with presence of prostate cancer, useful  
XX PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX PS Claim 1; Page 8784; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising  
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX CC specification or its complement. (I) is useful for: (a) assessing whether  
XX CC a patient is afflicted with prostate cancer; (b) monitoring the  
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy  
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX CC determining whether prostate cancer has metastasized in a patient; (h)  
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient  
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 459 BP; 120 A; 122 C; 111 G; 106 T; 0 U; 0 Other;  
SQ Sequence 459 BP; 120 A; 122 C; 111 G; 106 T; 0 U; 0 Other;  
Query Match 66.4%; Score 16.6; DB 5; Length 459;

Job time : 211.111 secs

Best Local Similarity 82.6%; Pred. No. 8.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTC 23  
 DB 82 AAAAAAGAGACACAGCGAGCTCCC 104

RESULT 50  
 ABV35393  
 ID ABV35393 standard; cDNA; 459 BP.  
 XX  
 AC ABV35393;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 35384.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW Pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 7374; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (f) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX  
 SQ Sequence 459 BP; 120 A; 122 C; 111 G; 106 T; 0 U; 0 Other;  
 Query Match 66.4%; Score 16.6; DB 5; Length 459;  
 Best Local Similarity 82.6%; Pred. No. 8.4e+02;  
 Mismatches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTC 23  
 DB 82 AAAAAAGAGACACAGCGAGCTCCC 104

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-15  
Perfect score: 25  
Sequence: 1 aaaaacgacacgagccctcgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	93.6	229	2	BB288882
2	23.4	93.6	297	6	CA895258
3	23.4	93.6	310	1	AW411607
4	23.4	93.6	320	2	BE630623
5	23.4	93.6	322	1	AA790812
6	23.4	93.6	326	8	W36874
7	23.4	93.6	328	1	AA871673
8	23.4	93.6	330	1	AW647013
9	23.4	93.6	330	5	BY497898
10	23.4	93.6	336	2	BE959526
11	23.4	93.6	337	6	CB570825
12	23.4	93.6	341	1	AA066781
13	23.4	93.6	342	1	AI843447
14	23.4	93.6	346	1	AA080458
15	23.4	93.6	348	2	BF319961
16	23.4	93.6	348	2	BE629477
17	23.4	93.6	352	2	BB685033
18	23.4	93.6	353	5	BY398944
19	23.4	93.6	354	5	BY497353
20	23.4	93.6	355	2	BE197629
21	23.4	93.6	357	3	BM899372
22	23.4	93.6	358	1	AW412420

C	23	23.4	93.6	361	2	BE994381	UI-M-B21-	BE994381	UI-M-B21-
24	23.4	93.6	361	5	BY608470	BY608470	BY608470	BY608470	BY608470
C	25	23.4	93.6	366	2	BE989916	UI-M-B21-	BE989916	UI-M-B21-
26	23.4	93.6	369	1	AA396989	AA396989	AA396989	AA396989	AA396989
27	23.4	93.6	370	3	BI655761	BI655761	603281362	603281362	603281362
28	23.4	93.6	370	5	BY504438	BY504438	BY504438	BY504438	BY504438
C	29	23.4	93.6	372	5	BF227139	BF227139	BF227139	BF227139
30	23.4	93.6	372	5	BY699023	BY699023	BY699023	BY699023	BY699023
31	23.4	93.6	374	1	AA396226	AA396226	VB44h01.r	VB44h01.r	VB44h01.r
32	23.4	93.6	375	2	BB744103	BB744103	BB744103	BB744103	BB744103
33	23.4	93.6	378	2	BB736988	BB736988	BB736988	BB736988	BB736988
C	34	23.4	93.6	379	2	BF459322	UI-M-B21-	BF459322	UI-M-B21-
35	23.4	93.6	380	1	AA250176	AA250176	mx85b12.r	mx85b12.r	mx85b12.r
C	36	23.4	93.6	381	1	AA088987	AA088987	mm51905.r	mm51905.r
37	23.4	93.6	381	1	AA240659	AA240659	mm25c02.r	mm25c02.r	mm25c02.r
38	23.4	93.6	382	5	BY737665	BB737665	BB737665	BB737665	BB737665
39	23.4	93.6	382	5	BY629510	BY629510	BY629510	BY629510	BY629510
40	23.4	93.6	382	5	BY685406	BY685406	BY685406	BY685406	BY685406
41	23.4	93.6	383	2	BF235341	BF235341	602026734	602026734	602026734
C	42	23.4	93.6	384	2	BE197726	BE197726	u877c09.x	u877c09.x
43	23.4	93.6	384	2	BE197726	BE197726	BY629024	BY629024	BY629024
44	23.4	93.6	384	5	BY629024	BY629024	BY699498	BY699498	BY699498
45	23.4	93.6	386	2	BF320359	u252h08.x	BF320359	u252h08.x	u252h08.x
46	23.4	93.6	386	5	BY642036	BY642036	BY642036	BY642036	BY642036
47	23.4	93.6	387	5	BY603881	BY603881	BY603881	BY603881	BY603881
C	48	23.4	93.6	387	5	BY630813	BY630813	BY630813	BY630813
49	23.4	93.6	387	5	BY660055	BY660055	BY660055	BY660055	BY660055
50	23.4	93.6	388	2	BB737836	BB737836	BB737836	BB737836	BB737836
51	23.4	93.6	389	5	BY401120	BY401120	BY401120	BY401120	BY401120
C	52	23.4	93.6	392	2	BE994338	UI-M-B21-	BE994338	UI-M-B21-
53	23.4	93.6	393	1	AA591164	AA591164	vn55908.r	vn55908.r	vn55908.r
54	23.4	93.6	393	5	BY427616	BY427616	BY427616	BY427616	BY427616
55	23.4	93.6	394	5	BY629324	BY629324	BY629324	BY629324	BY629324
56	23.4	93.6	395	5	BY438763	BY438763	BY438763	BY438763	BY438763
57	23.4	93.6	397	5	BY625494	BY625494	BY625494	BY625494	BY625494
58	23.4	93.6	397	5	BY627646	BY627646	BY627646	BY627646	BY627646
59	23.4	93.6	397	5	BY631643	BY631643	BY631643	BY631643	BY631643
60	23.4	93.6	398	5	BY575337	BY575337	BY575337	BY575337	BY575337
C	61	23.4	93.6	399	2	BE953304	UI-M-CD1-	BE953304	UI-M-CD1-
62	23.4	93.6	399	5	BY587583	BY587583	BY587583	BY587583	BY587583
C	63	23.4	93.6	399	6	CA893542	B0179G03-	CA893542	B0179G03-
64	23.4	93.6	400	3	BI794673	BI794673	IC63b07.x	IC63b07.x	IC63b07.x
65	23.4	93.6	400	5	BY579116	BY579116	BY579116	BY579116	BY579116
66	23.4	93.6	401	2	BB742737	BB742737	BB742737	BB742737	BB742737
C	67	23.4	93.6	401	7	CK330945	H8249F02-	CK330945	H8249F02-
68	23.4	93.6	402	3	QK265864	QK265864	NISC ff09	NISC ff09	NISC ff09
C	69	23.4	93.6	403	1	AW240772	uq33d08.x	AW240772	uq33d08.x
70	23.4	93.6	404	2	BE307297	BE307297	6010912222	6010912222	6010912222
71	23.4	93.6	404	5	BY601389	BY601389	BY601389	BY601389	BY601389
72	23.4	93.6	406	5	BY392307	BY392307	BY392307	BY392307	BY392307
73	23.4	93.6	406	5	BY638342	BY638342	BY638342	BY638342	BY638342
C	74	23.4	93.6	407	2	BE994642	UI-N-B21-	BE994642	UI-N-B21-
75	23.4	93.6	408	2	BB710131	BB710131	BB710131	BB710131	BB710131
76	23.4	93.6	408	5	BY511458	BY511458	BY511458	BY511458	BY511458
77	23.4	93.6	410	3	BI715778	BI715778	IC63b07.y	IC63b07.y	IC63b07.y
C	78	23.4	93.6	411	2	BF321381	u261f08.x	BF321381	u261f08.x
79	23.4	93.6	411	5	BY439640	BY439640	BY439640	BY439640	BY439640
80	23.4	93.6	411	5	BY574665	BY574665	BY574665	BY574665	BY574665
81	23.4	93.6	411	5	BY626670	BY626670	BY626670	BY626670	BY626670
C	82	23.4	93.6	413	2	BF457487	UI-N-B21-	BF457487	UI-N-B21-
83	23.4	93.6	413	2	BB743431	BB743431	BB743431	BB743431	BB743431
C	84	23.4	93.6	414	2	BO618133	L0951G04-	BO618133	L0951G04-
85	23.4	93.6	414	2	BB677028	BB677028	BB677028	BB677028	BB677028
86	23.4	93.6	414	2	BB686470	BB686470	BB686470	BB686470	BB686470
C	87	23.4	93.6	414	3	BM203581	C0256F03-	BM203581	C0256F03-
88	23.4	93.6	414	5	BY649734	BY649734	BY649734	BY649734	BY649734
89	23.4	93.6	415	1	BY502341	BY502341	BY502341	BY502341	BY502341
C	90	23.4	93.6	416	1	AW763873	ur66h09.x	AW763873	ur66h09.x
91	23.4	93.6	416	2	BB828941	BB828941	BB828941	BB828941	BB828941
92	23.4	93.6	417	1	AA079500	AA079500	AA079500	AA079500	AA079500
93	23.4	93.6	420	2	BB742412	BB742412	BB742412	BB742412	BB742412
94	23.4	93.6	421	2	BB677378	BB677378	BB677378	BB677378	BB677378
95	23.4	93.6	421	5	BY601791	BY601791	BY601791	BY601791	BY601791

BE994381	UI-M-BZ1-
BE608470	BY608470
BE989916	UI-M-BZ1-
AA396989	mx86c05.f
BI655761	603281362
BY504438	BY504438
BF227139	uz20805.x
BY699023	BY699023
AA396226	VB44H01.f
BB744103	BB744103
BB736988	BB736988
BF459322	UI-M-BZ1-
AA250176	mx85b12.f
AA088987	mm51905.f
AA240659	mv22602.f
BY629510	BY629510
BY685406	BY685406
BF235341	602026734
BE197726	u877c09.x
BY629024	BY629024
BY699498	BY699498
BF320359	uz25208.x
BY642036	BY642036
BY603881	BI603881
BY630813	BY630813
BY660055	BY660055
BB737836	BB737836
BY401120	BY401120
BE994338	UI-M-BZ1-
AA591164	vn53908.f
BY427616	BY427616
BY629324	BY629324
BY438763	BY438763
BY625494	BY625494
BY627646	BI627646
BY631643	BY631643
BY575337	BY575337
BE953304	UI-M-CD1-
BY587583	BY587583
CA893542	B0179603-
BI794673	1c63b07.x
BY579116	BY579116
BB742737	BB742737
CK330945	H8249F02-
QK265864	NTSC_ff09
AW240772	uq33d08.x
BE307297	601091222
BY601389	BY601389
BY392307	BY392307
BY638342	BY638342
BE994642	UI-M-BZ1-
BB710131	BB710131
BY511458	BY511458
BI715778	1c63b07.y
BF321381	uz61f08.x
BY439640	BY439640
BY574665	BY574665
BY626670	BY626670
BF457487	UI-M-BZ1-
BB743431	BB743431
BO618133	LO951G04-
BB677028	BB677028
BB686470	BB686470
BM203581	CO256F03-
BY649734	BY649734
BY502341	BY502341
AW763873	uz66h09.x
BB828941	BB828941
AA079500	zm96h10.s
BB742412	BB742412
BB677378	BB677378
BY601791	BY601791

96	23.4	93.6	422	2	BG148427	BG148427 uu78a02.y
97	23.4	93.6	422	2	BB743760	BB743760 BB743760.y
98	23.4	93.6	422	2	BB135363	BB135363 ug27a09.y
99	23.4	93.6	422	3	BI853653	BI853653 603380781
100	23.4	93.6	423	2	BB742024	BB742024 BB742024
101	23.4	93.6	423	2	BB780250	BB780250 BB780250
102	23.4	93.6	423	2	BB743140	BB743140 BB743140
103	23.4	93.6	424	3	BI665580	BI665580 603287857
104	23.4	93.6	425	2	BB781758	BB781758 BB781758
c 105	23.4	93.6	425	3	BM234099	BM234099 K0345C04-
106	23.4	93.6	425	5	BY697991	BY697991 BY697991
c 107	23.4	93.6	426	1	AI195797	AI195797 ui50gl1.x
c 108	23.4	93.6	426	2	BF322726	BF322726 ma332d01.
109	23.4	93.6	426	2	BB676021	BB676021 BB676021
110	23.4	93.6	426	2	BF011812	BF011812 us37a04.y
111	23.4	93.6	427	2	BB668430	BB668430 BB668430
112	23.4	93.6	427	2	BB778573	BB778573 BB778573
c 113	23.4	93.6	427	5	BY394075	BY394075 BY394075
c 114	23.4	93.6	428	1	AW681738	AW681738 EST00456
c 115	23.4	93.6	428	1	AA590734	AA590734 vml7a11.r
116	23.4	93.6	428	2	BB778575	BB778575 BB778575
117	23.4	93.6	428	2	BB789474	BB789474 BB789474
c 118	23.4	93.6	430	2	BB835361	BB835361 BB835361
c 119	23.4	93.6	431	2	BG064583	BG064583 H3021E03-
c 120	23.4	93.6	431	2	BB673525	BB673525 BB673525
121	23.4	93.6	431	5	BY567124	BY567124 BY567124
c 122	23.4	93.6	432	2	BF226725	BF226725 uz15a09.x
123	23.4	93.6	432	8	WA9123	WA9123 mc91b12.r1
124	23.4	93.6	434	5	BY633095	BY633095 BY633095
125	23.4	93.6	437	1	AA289650	AA289650 vb2da03.r
126	23.4	93.6	437	5	BY443781	BY443781 BY443781
127	23.4	93.6	437	8	W83242	W83242 mf23g07.r1
c 128	23.4	93.6	438	1	AA275899	AA275899 vc35f09.r
c 129	23.4	93.6	438	1	AW987304	AW987304 uf16e08.x
130	23.4	93.6	438	2	BB781720	BB781720 BB781720
131	23.4	93.6	438	2	BB833996	BB833996 BB833996
132	23.4	93.6	438	5	BY644932	BY644932 BY644932
c 133	23.4	93.6	440	1	AI461748	AI461748 ub78e10.x
134	23.4	93.6	440	8	W29761	W29761 mc07g05.r1
135	23.4	93.6	441	2	BG971197	BG971197 602839766
136	23.4	93.6	442	5	BY387653	BY387653 BY387653
137	23.4	93.6	445	2	BB684314	BB684314 BB684314
138	23.4	93.6	448	1	AA273478	AA273478 vc37c07.r
c 139	23.4	93.6	449	2	BB783062	BB783062 BB783062
c 140	23.4	93.6	450	3	BQ265133	BQ265133 NISC if03
c 141	23.4	93.6	451	2	BF320876	BF320876 uz56h03.x
142	23.4	93.6	451	5	BY393605	BY393605 BY393605
c 143	23.4	93.6	451	6	CA885726	CA885726 B0122B01-
c 144	23.4	93.6	452	1	AW701298	AW701298 un81c09.x
c 145	23.4	93.6	452	2	BF147639	BF147639 us37a04.x
c 146	23.4	93.6	453	1	AV021619	AV021619 AV021619
c 147	23.4	93.6	453	2	BG141550	BG141550 ia90a06.x
148	23.4	93.6	453	2	BB731974	BB731974 BB731974
149	23.4	93.6	453	2	BB782938	BB782938 BB782938
150	23.4	93.6	454	1	AA097672	AA097672 mn85c05.r

1 (bases 1 to 229)

Konno H., Aizawa, K., Akahira S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagasaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tonaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

Location/Qualifiers  
1. .229  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="B020034112"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATCTCGAGTGAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

Query Match 93.6%; Score 23.4; DB 2; Length 229;  
Best Local Similarity 96.0%; Pred. No. 5.7;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGACGAGCCCTCGG 25  
|||||  
DB 88 AAAAAACGACACTCGGAGCCCTCGG 112

ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 229;  
Best Local Similarity 96.0%; Pred. No. 5.7;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGACGAGCCCTCGG 25  
|||||  
DB 88 AAAAAACGACACTCGGAGCCCTCGG 112

FEATURES

source

1. .229  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="B020034112"  
/tissue\_type="egg"  
/dev\_stage="2 cells"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 2 cells egg"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATCTCGAGTGAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ALIGNMENTS

229 bp mRNA linear EST 09-JUL-2000  
BB288882 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA  
clone B020034112 3', mRNA sequence.

BB288882 229 bp mRNA linear EST 09-JUL-2000  
BB288882 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA  
clone B020034112 3', mRNA sequence.

BB288882.1 GI:8989331  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.





REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 320)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Other\_ESTs: uu38d12.y1  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb-r@mail.nih.gov](mailto:cgapsb-r@mail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:1083835.

FEATURES  
sourceLocation/Qualifiers  
1. .320

/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3374231"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/clone\_lib="Soares mammary gland NMLMG"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 320;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
Db 318 AAAAAACGACACTGCGGCCCTCGG 294  
|||||

RESULT 5  
AA790812 322 bp mRNA linear EST 06-FEB-1998  
LOCUS mw19d12.r1 Soares mammary\_gland\_NbMMG Mus musculus cDNA clone  
DEFINITION IMAGE:1244279 5', mRNA sequence.

ACCESSION AA790812  
VERSION AA790812.1 GI:2850932  
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 322)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:657967

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 311.

FEATURES  
sourceLocation/Qualifiers  
1. .322

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1244279"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares mammary\_gland NbMMG"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGCGCGAATGTTTGTGTGTGTGTGTGTGTGT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru KO, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 322;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
Db 114 AAAAAACGACACTGCGGCCCTCGG 138  
|||||

## RESULT 6

W36874

LOCUS

DEFINITION

W36874.1 Soares mouse p3NM19.5 Mus musculus cDNA clone

IMAGE:335443 5', mRNA sequence.

ACCESSION

W36874

VERSION

W36874.1 GI:1318149

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 326)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

High quality sequence stop: 321.

Seq primer: ETPrimer

Location/Qualifiers

1. .326

/organism="Mus musculus"

/mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:335443"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NM19.5"  
 /note="Vector: pT73D (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTGAAGTCGAGCGCGCATTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 8; Length 326;  
 Best Local Similarity 96.0%; Pred. No. 5.8;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCGG 25  
 |||||  
 DB 290 AAAAAACGACACTGCGAGCCCTCGG 314

## RESULT 7

AA871673  
 LOCUS  
 DEFINITION  
 vq39e03.r1 Barstead bowel MPLRB9 Mus musculus cDNA clone  
 IMAGE:1096636 5', mRNA sequence.  
 AA871673  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 328)

## REFERENCE

AUTHORS  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Spector, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

## TITLE

JOURNAL  
 COMMENT  
 The WashU-HHMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of MedicineP  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:602868

Putative full length read  
 vector to vector length is 517  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 297.

## FEATURES

source  
 1..328  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1096636"  
 /tissue\_type="bowel"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"

/clone\_lib="Barstead bowel MPLRB9"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTGAAGTCGAGCGCGCATTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adapters  
 (Pharmacia), digested with Not I and cloned into the  
 Not I and Eco RI sites of the modified pT73 vector.  
 Source irradiated bowel harvested 72 hours after  
 irradiation (1400 Gys). Library constructed by Bob  
 Barstead."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 328;  
 Best Local Similarity 96.0%; Pred. No. 5.8;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCGG 25  
 |||||  
 DB 197 AAAAAACGACACTGCGAGCCCTCGG 221

## RESULT 8

AW647013  
 LOCUS  
 DEFINITION  
 EST325603 R3DA Mus musculus cDNA clone R3DAB40, mRNA sequence.  
 AW647013  
 ACCESSION  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 330)

## REFERENCE

AUTHORS  
 Earle-Hughes, J., Cho, J., Hansen, T.S., Lee, H.H., Quackenbush, J.,  
 Adams, M.D., Fraser, C.M. and Venter, J.C.  
 Generation of ESTs from Murine adipose, differentiated 3T3 cell  
 line

## TITLE

JOURNAL  
 COMMENT  
 Unpublished (1994)  
 Other ESTs: TC104313  
 Contact: Julie Earle-Hughes  
 TIGR

The Institute For Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

## FEATURES

source  
 1..330  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /cultivar="3T3"  
 /db\_xref="taxon:10090"  
 /clone="R3DAB40"  
 /tissue\_type="adipose"  
 /clone\_lib="R3DA"  
 /note="Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 330;  
 Best Local Similarity 96.0%; Pred. No. 5.8;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCGG 25  
 |||||  
 DB 248 AAAAAACGACACTGCGAGCCCTCGG 272

## RESULT 9

BY497898  
 LOCUS  
 DEFINITION  
 BY497898 RIKEN full-length enriched, bone marrow macrophage Mus  
 musculus cDNA clone 183052D10 3', mRNA sequence.

ACCESSION	BY497898	FEATURES	source
VERSION	BY497898.1	Location/Qualifiers	1..330
KEYWORDS	EST.	organism="Mus musculus"	
SOURCE	Mus musculus (house mouse)	/mol_type="mRNA"	
ORGANISM		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/clone="I830052D10"	
		/tissue_type="bone marrow"	
		/cell_type="macrophage"	
		/clone_lib="RIKEN full-length enriched, bone marrow macrophage"	
REFERENCE			
AUTHORS			
		Query Match	93.6%; Score 23.4; DB 5; Length 330;
		Best Local Similarity	96.0%; Pred. No. 5.8;
		Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
		QY	1 AAAAAACGACACGAGCCCTCGG 25
		DB	17 AAAAAACGACACTCGGACCTCGG 41
		RESULT 10	
		BE995926/c	
		LOCUS	336 bp mRNA linear EST 05-OCT-2000
		DEFINITION	UI-M-B21-biw-b-03-0-UI.s1 NIH BMAP MH12.s1 Mus musculus cDNA clone
			UI-M-B21-biw-b-03-0-UI 3', mRNA sequence.
		ACCESSION	BE995926
		VERSION	BE995926.1
		KEYWORDS	GI:10680664
		SOURCE	EST.
		ORGANISM	Mus musculus (house mouse)
			Mus musculus
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
			Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
			Sciurognathi; Muridea; Muridae; Murinae; Mus.
		REFERENCE	1 (bases 1 to 336)
		AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
		TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
		JOURNAL	Genome Res. 6 (9), 791-806 (1996)
		PUBMED	8889548
		COMMENT	Contact: Chin, H National Institute of Mental Health 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA Tel: 301 443 1706 Fax: 301 443 9890 Email: mES@mail.nih.gov The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 1-53, >AT rich#Low complexity Seq primer: M13 Forward POLYA=Yes.
		FEATURES	Location/Qualifiers
		source	1..336
			organism="Mus musculus"
			/mol_type="mRNA"
			/strain="C57BL/6J"
			/db_xref="taxon:10090"
			/clone="UI-M-B21-biw-b-03-0-UI"

BY497898  
BY497898.1 GI:26932277  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 330)  
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,  
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,  
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,  
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,  
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,  
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,  
Guernicich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,  
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,  
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G.,  
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,  
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,  
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,  
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,  
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,  
Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,  
Sakazume,N., Sato,K., Shira,T., Waki,K., Kawai,J., Aizawa,K.,  
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC) Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.jp, URL:http://genome.gsc.riken.jp/  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by David A. Hume ( Depts. of Biochemistry  
and Microbiology/Parasitology Institute for Molecular Bioscience  
University of Queensland Brisbane, Q 4072 Australia ) whose

```

/dev stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_MHI2_S1"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The
NIH_BMAP_MHI2_S1 library is a substracted library derived
from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_TISSUE=hippocampus
TAG_LIB=NIH_BMAP_MHI2_S1
TAG_SEQ=TAGCC"

```

## ORIGIN

```

Query Match      93.6%; Score 23.4; DB 2; Length 336;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAACGACACGAGCCCTCGG 25
    |||||
Db 306 AAAAAACGACACTCGAGCCCTCGG 282
    |||||

```

```

RESULT 11
CB570825 LOCUS
DEFINITION AGENCOURT 12976722 NIH_MGC_165 Mus musculus cDNA clone
IMAGE:30278425 5', mRNA sequence.
CB570825
CB570825.1 GI:29490355
EST.
Mus musculus (house mouse)

```

## ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 337)
TITLE NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcapbe-remail.nih.gov

```

```

Tissue Procurement: Dr. Leslie L. Heckert
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM32 row: n column: 02
High quality sequence stop: 337.

```

## FEATURES

```

source
1..337
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30278425"
/tissue_type="primary cultures of Sertoli cells"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_165"
/notes="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI
(ggcatattggcc); Site 2: SfiI (ggcgctctggcc); 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4 kb
(range 0.6-3.5 kb). 15/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH_MGC Library."

```

## ORIGIN

```

Query Match      93.6%; Score 23.4; DB 6; Length 337;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCCCTCGG 25
    |||||
Db 22 AAAAAACGACACTCGAGCCCTCGG 46
    |||||

```

## RESULT 12

```

AA066781 LOCUS
DEFINITION mm16e06.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
clone IMAGE:521698 5', mRNA sequence.
AA066781
AA066781.1 GI:1564548
EST.
Mus musculus (house mouse)

```

## KEYWORDS

## SOURCE

## ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

## REFERENCE

## AUTHORS

## TITILE

## JOURNAL

## COMMENT

## Unpublished (1996)

## Contact: Marra M/Mouse EST Project

## WashU-HHMI Mouse EST Project

## Washington University School of MedicineP

## 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

## Tel: 314 286 1800

## Fax: 314 286 1810

## Email: mouseest@watson.wustl.edu

## This clone is available royalty-free through LLNL; contact the

## IMAGE Consortium (info@image.llnl.gov) for further information.

## MGI:315546

## Seq primer: -28ml3 rev1 ET from Amersham

## High quality sequence stop: 300.

## FEATURES

## source

## 1..341

## /organism="Mus musculus"

## /mol\_type="mRNA"

## /db\_xref="taxon:10090"

## /clone="IMAGE:521698"

## /tissue\_type="diaphragm"

## /dev\_stage="adult"

## /lab\_host="SOLR (kanamycin resistant)"

## /clone\_lib="Stratagene mouse diaphragm (#937303)"

## /notes="Organ: diaphragm; Vector: pBluescript SK-; Site 1:

## EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA

## prepared from diaphragm muscle. Primer: Oligo dt. Average

## insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor

## sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

## CTCGAGTTTTTTTTTTTTTTT 3"

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 24; Conservative

## 0; Mismatches

## 1; Indels

## 0; Gaps

## 0;

## Qy

## 1 AAAAAACGACACGAGCCCTCGG 25

## |||||

## Db

## 274 AAAAAACGACACTCGAGCCCTCGG 298

## |||||

## RESULT 13

## AI843447/c

## LOCUS

## DEFINITION

## UI-M-AQ1-aec-d-11-0-UI.s1 NIH\_BMAP\_MHI\_N Mus musculus cDNA clone

## EST 14-JUL-1999

```

UI-M-AQ1-aec-d-11-0-UI 3', mRNA sequence.
ACCESSION AI843447
VERSION AI843447.1 GI:5477660
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 342)
REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643 USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized hippocampus library cDNA library Preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distributing of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
The following repetitive elements were found in this cDNA sequence:
1-73, >AT rich#Low complexity
Seq primer: M13 Forward
POLYA=yes.
FEATURES
source
1..342
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AQ1-aec-d-11-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MHI_n library is a normalized library constructed
from mouse hippocampus. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Miller Laboratories.
TAG_TISSUE=hippocampus
TAG_LIB=NIH_BMAP_MHI_N
TAG_SEQ=TTCGA"
ORIGIN
Query Match 93.6%; Score 23.4; DB 1; Length 342;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAACGACACGAGCGCCCTCGG 25
|||||
Db 326 AAAAAACGACACTGCGAGCCCTCGG 302
|||||
RESULT 14
AA080458 346 bp mRNA linear EST 15-FEB-1997
LOCUS mn87g10.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
DEFINITION

```

```

IMAGE:5511106 5', mRNA sequence.
ACCESSION AA080458
VERSION AA080458.1 GI:1619464
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 346)
REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:331898
Seq primer: -28ml3 rev1 ET from Amer sham
High quality sequence stop: 324.
Location/Qualifiers
1..346
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:551106"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse Tcell 937311"
/notes="Organ: blood; Vector: pBluescript SK-; Site_1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGAG
3' ~3' adaptor sequence: 5' CTCGAGTTTITTTTTTTT 3'"
ORIGIN
Query Match 93.6%; Score 23.4; DB 1; Length 346;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAACGACACGAGCGCCCTCGG 25
|||||
Db 257 AAAAAACGACACTGCGAGCCCTCGG 233
|||||
RESULT 15
BF319961/c 346 bp mRNA linear EST 29-DEC-2000
LOCUS uz44g01.x1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3671952 3',
DEFINITION mRNA sequence.
ACCESSION BF319961
VERSION BF319961.1 GI:11268813
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 346)
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```

## COMMENT

Other ESTs: uz44g01.y1  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Cloning Distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 image.llnl.gov/image/html/iresources.shtml

MGI:1432720

High quality sequence stop: 325.

## FEATURES

Location/Qualifiers

## source

1..346

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3671952"

/sex="female, virgin"

/tissue\_type="infiltrating ductal carcinoma"

/dev\_stage="5 months"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP Mam6"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

## ORIGIN

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 2; Length 346;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 AAAAAACGACACGAGCCCTCGG 25

|||||

321 AAAAAACGACACTCGAGCCCTCGG 297

|||||

## RESULT 16

BE629477

LOCUS

DEFINITION BE629477 348 bp mRNA linear EST 25-AUG-2000

IMAGE:3374231 5', mRNA sequence.

ACCSSION BE629477

VERSION BE629477.1 GI:9912165

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 348)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1083835

Seq primer: -40RP from Gibco

High quality sequence stop: 311.

Location/Qualifiers

1..348

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:3374231"

/sex="female (lactating)"

/tissue\_type="mammary gland"

## FEATURES

## source

1..348

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:3374231"

/sex="female (lactating)"

/tissue\_type="mammary gland"

/lab\_host="DH10B"

/clone\_lib="Soares mammary gland NMLMG"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 348;

Best Local Similarity 96.0%; Pred. No. 5.8;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 AAAAAACGACACGAGCCCTCGG 25

|||||

36 AAAAAACGACACTCGAGCCCTCGG 60

|||||

## RESULT 17

BE685033

LOCUS

DEFINITION BE685033 352 bp mRNA linear EST 10-OCT-2001

murine duc Mus musculus cDNA clone 6820416K12 3', mRNA

sequence.

ACCSSION BE685033

VERSION BE685033.1 GI:16011766

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 352)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,

Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,

Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,

Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,

Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,

Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,

Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.

and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,

Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)







BY497353.1	GI:26831732
EST.	
KEYWORDS	Mus musculus (house mouse)
SOURCE	Mus musculus
ORGANISM	Mus musculus
REFERENCE	
AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,K., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalia,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Giassi,C., Godik,A., Gough,J., Grimmerond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierzki,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/ Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by David A. Hume ( Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia ) whose assistance we gratefully acknowledge.

FEATURES	source
1. 354	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/clone="I830049H02"
	/tissue_type="bone marrow"
	/cell_type="macrophage"
	/clone_lib="RIKEN full-length enriched, bone marrow macrophage"
ORIGIN	
Query Match	93.6%; Score 23.4; DB 5; Length 354;
Best Local Similarity	96.0%; Pred. No. 5.8;
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 AAAAAACGACACAGCGAGCCCTCGG 25 
Db	40 AAAAAACGACATGCGAGCCCTCGG 64 
RESULT 20	
BE197629/c	
LOCUS	BE197629.1 GI:8709798
DEFINITION	Mus musculus (house mouse)
ACCESSION	BE197629
VERSION	EST.
KEYWORDS	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
ORGANISM	1 (bases 1 to 355)
REFERENCE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
AUTHORS	Unpublished (1997)
TITLE	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps@mail.nih.gov">cgaps@mail.nih.gov</a> Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov/image/html/iresources.shtml">image.llnl.gov/image/html/iresources.shtml</a>
JOURNAL	
COMMENT	
MG1:1068426	
Possible reversed clone:	polyT not found
Seq primer:	-40UP from Gibco
High quality sequence stop:	347.
Location/Qualifiers	
1. 355	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="NMRI"
	/db_xref="taxon:10090"
	/clone="IMAGE:3257062"
	/tissue_type="tumor, gross tissue"
	/dev_stage="5 months"
	/lab_host="DH10B"
	/clone_lib="NCI CGAP Mam4"
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth

ORIGIN	and Differentiation 7, 3-11 (1996)." TAG_TISSUE=subfornical organ and postrema TAG_LIB=UI-M-DJ1 TAG_SEQ=GCTACATGAT"	
	Query Match 93.6%; Score 23.4; DB 2; Length 355; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AAAAAACGACACAGCGAGCCCTCGG 25 	
Db	288 AAAAAACGACACTGCGAGCCCTCGG 264 	
RESULT 21	BM899372/c	
LOCUS	UI-M-DJ1-btq-c-05-0-UI.s1 NIH BMAP DJ1 Mus musculus cDNA clone	
DEFINITION	UI-M-DJ1-btq-c-05-0-UI 3', mRNA sequence.	
ACCESSION	BM899372	
VERSION	BM899372.1 GI:19382160	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.	
	1 (bases 1 to 357)	
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.	
	Normalization and subtraction: two approaches to facilitate gene discovery	
TITLE	Genome Res. 6 (9), 791-806 (1996)	
JOURNAL	8889548	
PUBMED	Contact: Chin, H	
COMMENT	National Institute of Mental Health 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA Tel: 301 443 1706 Fax: 301 443 9890 Email: m85@mail.nih.gov Tissue Procurement: Dr. Robin Davison CDNA Library preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 1-21, >AT rich#Low_complexity (matched complement) Seq primer: M13 FORWARD POLYA=Yes.	
FEATURES	Location/Qualifiers	
	1..357 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="UI-M-DJ1-btq-c-05-0-UI" /tissue_type="subfornical organ and postrema" /dev_stage="Adult" /lab_host="DH10B (Life Technologies) (T1 phage resistant)" /clone_lib="NIH BMAP DJ1" /note="Organ: brain; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site1: EcoR I; Site 2: Not I; NIH BMAP DJ1 is a cDNA library containing the following tissues: subfornical organ and postrema. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTACATGAT."	
ORIGIN	Query Match 93.6%; Score 23.4; DB 1; Length 358; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
	QY 1 AAAAAACGACACAGCGAGCCCTCGG 25 	
Db	325 AAAAAACGACACTGCGAGCCCTCGG 301 	
RESULT 22	AW412420	
LOCUS	uq41b09.x1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:2811929 3', mRNA sequence.	
DEFINITION	AW412420	
ACCESSION	AW412420	
VERSION	AW412420.1 GI:6938293	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.	
	1 (bases 1 to 358)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
TITLE	Unpublished (1997)	
JOURNAL	Other ESTs: uq41b09.y1	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys CDNA Library preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html	
FEATURES	Location/Qualifiers	
	1..358 /organism="Mus musculus" /mol_type="mRNA" /strain="mix FVB/N, C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:2811929" /tissue_type="tumor, gross tissue" /dev_stage="7 months" /lab_host="DH10B" /clone_lib="NCI CGAP Mam5" /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"	
ORIGIN	Query Match 93.6%; Score 23.4; DB 1; Length 358; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
	QY 1 AAAAAACGACACAGCGAGCCCTCGG 25 	
Db	325 AAAAAACGACACTGCGAGCCCTCGG 301 	

RESULT 23	BE994381	361 bp	mRNA	linear	EST 05-OCT-2000
LOCUS	UI-M-BZ1-bjm-e-10-0-UI.s1	NIH BMAP MH12 S1	Mus musculus	cDNA clone	EST 15-DEC-2002
DEFINITION	UI-M-BZ1-bjm-e-10-0-UI 3', mRNA sequence.				
ACCESSION	BE994381				
VERSION	BE994381.1	GI:10677563			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 361)				
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery				
PUBLISHED	Genome Res. 6 (9), 791-806 (1996)				
COMMENT	8889548 Contact: Chin, H National Institute of Mental Health 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA Tel: 301 443 1706 Fax: 301 443 9890 Email: mstetm@nih.gov The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements Seq primer: M13 Forward POLYA=Yes.				
FEATURES	source	1..361	Location/Qualifiers		
		/organism="Mus musculus"			
		/mol_type="mRNA"			
		/db_xref="taxon:10090"			
		/clone="UI-M-BZ1-bjm-e-10-0-UI"			
		/dev_stage="27-32 days"			
		/lab_host="DH10B (Life Technologies)"			
		/clone_lib="NIH BMAP MH12 S1"			
		/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BMAP MH12 S1 library is a subtracted library derived from NIH BMAP MH12. NIH BMAP MH12 is a library derived from mouse hippocampus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu."			
		TAG TISSUE=hippocampus			
		TAG LIB=NIH BMAP MH12 S1			
		TAG_SEQ=TAGTC			
ORIGIN					
Query Match	93.6%	Score 23.4;	DB 2;	Length 361;	
Best Local Similarity	96.0%	Pred. NO. 5.8;			
Matches	24;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	AAAAAACGACACGAGCCCTCGG 25			
Db	333	AAAAAACGACACGAGCCCTCGG 309			
RESULT 24	BY608470	361 bp	mRNA	linear	EST 15-DEC-2002
LOCUS	BY608470	RIKEN full-length enriched, visual cortex Mus musculus			
DEFINITION	cDNA clone K230302L24 3', mRNA sequence.				
ACCESSION	BY608470.1	GI:26943652			
VERSION	EST.				
KEYWORDS	Mus musculus (house mouse)				
SOURCE	Mus musculus				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 361)				
TITLE	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Niikado,I., Oato,N., Saito,R., Suzuki,H., Yamana,K., I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.				
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
	Nature 420, 563-573 (2002)				
CONTACT	Yoshihide Hayashizaki				
	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute				
	The Institute of Physical and Chemical Research (RIKEN)				
	1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
	Tel: 81-45-503-9222				
	Fax: 81-45-503-9216				
	Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/				
	Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission				
	Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)				
	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)				
	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)				
	Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)				
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				



ORIGIN	Bonaldo."	musculus cDNA clone I830085H24 3', mRNA sequence.
Query Match	93.6%; Score 23.4; DB 1; Length 369;	
Best Local Similarity	96.0%; Pred. No. 5.8;	
Matches 24; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AAAAAACGACACAGCGAGCCCTCGG 25	
DB	256 AAAAAACGACACTGCGAGCCCTCGG 280	
RESULT 27		
BI655761		
LOCUS	370 bp mRNA linear EST 12-SRP-2001	
DEFINITION	603281362F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5325831 5', mRNA sequence.	
ACCESSION	BI655761	
VERSION	BI655761.1 GI:155669997	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 370)	
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM11826 row: k column: 16 High quality sequence stop: 370. Location/Qualifiers 1..370 /organism="Mus musculus" /mol_type="mRNA" /strain="NMRI" /db_xref="taxon:10090" /clone="IMAGE:5325831" /tissue_type="tumor, gross tissue" /dev_stage="5 months" /lab_host="DH10B" /clone_lib="NCI_CGAP Mam4" /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali; Site 2: Not1; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."	
FEATURES		
source		
ORIGIN		
Query Match	93.6%; Score 23.4; DB 3; Length 370;	
Best Local Similarity	96.0%; Pred. No. 5.8;	
Matches 24; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AAAAAACGACACAGCGAGCCCTCGG 25	
DB	45 AAAAAACGACACTGCGAGCCCTCGG 69	
RESULT 28		
BI504438		
LOCUS	370 bp mRNA linear EST 14-DEC-2002	
DEFINITION	RIKEN full-length enriched, bone marrow macrophage Mus	

ACCESSION	BI504438	
VERSION	BI504438.1 GI:26838817	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 370)	
JOURNAL	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gibsi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazawa, T., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL	Nature 420, 563-573 (2002)	
PUBLISHED	12466851	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by David A. Hume ( Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience	

University of Queensland Brisbane, Q 4072 Australia ) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

source  
1. .370  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="I830085H24"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_lib="RIKEN full-length enriched, bone marrow  
macrophage"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 5; Length 370;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
45 AAAAAACGACACTGCGAGCCCTCGG 69

## RESULT 29

### BF227139/c

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

BF227139 372 bp mRNA linear EST 29-DRC-2000  
uz20b05.x1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3669585 3',  
mRNA sequence.  
BF227139  
EST.  
1 (bases 1 to 372)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Other ESTs: uz20b05.y1  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

MG1:1430353

High quality sequence stop: 366.

# FEATURES

source  
1. .372  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="mix FVB/N, CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3669585"  
/tissue\_type="tumox, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam5"  
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Salt;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 372;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
316 AAAAAACGACACTGCGAGCCCTCGG 292

## RESULT 30

### BY699023

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

BY699023 372 bp mRNA linear EST 16-DEC-2002  
BY699023 RIKEN full-length enriched, osteoclast-like cell Mus  
musculus cDNA clone I420114J10 3', mRNA sequence.  
BY699023  
EST.  
1 (bases 1 to 372)  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

## REFERENCE

### AUTHORS

1 (bases 1 to 372)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gibsi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawachi, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Sasaki, D., Sato, K., Shibata, K.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

## TITLE

### JOURNAL

### PUBMED

### COMMENT

12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@gsc.riken.jp](mailto:genome-res@gsc.riken.jp), [URL:http://genome.gsc.riken.jp/](mailto:URL:http://genome.gsc.riken.jp/)  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tsgami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to



prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Takashi Ishikawa ( Department of Surgery 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan ) whose assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source  
 1. .372  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="I420114J10"  
 /cell\_type="osteoclast-like cell"  
 /clone\_lib="RIKEN full-length enriched, osteoclast-like cell"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 5; Length 372;  
 Best Local Similarity 96.0%; Pred. No. 5.8;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25

Db 59 AAAAAACGACACGAGCGCCCTCGG 83

## RESULT 31

AA396226

LOCUS

DEFINITION v044h01.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone IMAGE.751825 5', mRNA sequence. EST 25-APR-1997

ACCESSION AA396226

VERSION 1 GI:2049249

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS

1 (bases 1 to 374)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilton, R. and Waterston, R.

TITLE

JOURNAL

COMMENT

The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of MedicineP  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: [mouseest@wustl.edu](mailto:mouseest@wustl.edu)  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:460809

Seq primer: -28m13 rev2 ET from Amersham.

## FEATURES

.. source

Location/Qualifiers

1. .374

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:751825"  
 /sex="male"  
 /tissue\_type="lymph node"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares mouse lymph node NbMLN"

/notes="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site 2: Eco RI; ist strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTCACTGAGCGCGCGATCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 374;  
 Best Local Similarity 96.0%; Pred. No. 5.8;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25

Db 111 AAAAAACGACACGAGCGCCCTCGG 135

## RESULT 32

BB744103

LOCUS

DEFINITION BB744103 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F520015N21 3', mRNA sequence. EST 15-OCT-2001

ACCESSION BB744103

VERSION BB744103.1

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS

1 (bases 1 to 375)  
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watanabe, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL

COMMENT

Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9226

Email: [genome-res@sc.riken.jp](mailto:genome-res@sc.riken.jp), URL: <http://genome.gsc.riken.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.





20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mES@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the hippocampus tissue cDNA library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements. The following repetitive elements were found in this cDNA sequence: 1-21,  
>AT rich#Low complexity  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

source

Location/Qualifiers

```
1. .379
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BZ1-bmc-a-08-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_MHI2_S1"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MHI2_S1 library is a subtracted library derived
from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_TISSUE=hippocampus
TAG_LIB=NIH_BMAP_MHI2_S1
TAG_SEQ=TAGTC"
```

## ORIGIN

```
Query Match 93.6%; Score 23.4; DB 2; Length 379;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25
|||||
Db 334 AAAAAACGACACTGCGAGCCCTCGG 310
```

## RESULT 35

AA250176

LOCUS

```
DEFINITION mx85b12.r1 Soares mouse NML BMAP musculus cDNA clone IMAGE:693119
5', mRNA sequence.
```

ACCESSION

AA250176

VERSION

AA250176.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 380)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

TITLE

The WashU-HHMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:426679

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 379.

## FEATURES

source

Location/Qualifiers

```
1. .380
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:693119"
/tissue_type="Liver"
/lab_host="DH10B"
/clone_lib="Soares mouse NML"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAAGTGGACGGCCGGAATCTTTTTCCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
```

## ORIGIN

```
Query Match 93.6%; Score 23.4; DB 1; Length 380;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25
|||||
Db 268 AAAAAACGACACTGCGAGCCCTCGG 292
```

## RESULT 36

AA088987/c

LOCUS

```
DEFINITION mm51905.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
clone IMAGE:525080 5', mRNA sequence.
```

ACCESSION

AA088987

VERSION

AA088987.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 381)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:318928

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 325.

Location/Qualifiers

```

source
1. .381
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:525080"
/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse melanoma (#937312)"
Note="Organ: skin; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. From M2 cells, a highly metastatic derivative of
the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Query Match 93.6%; Score 23.4; DB 1; Length 381;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25
|||||
Db 234 AAAAAACGACACTGCGAGCCCTCGG 258
|||||

RESULT 38
BB737665
LOCUS
DEFINITION
BB737665 382 bp mRNA linear EST 15-OCT-2001
musculus cDNA clone F430014L12 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 382)
AUTHORS
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tonaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1. .382
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

FEATURES
source
1. .381
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:655778"
/tissue_type="kidney"
/dev_stage="newborn (day 0)"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="GuayWoodford Beier mouse kidney day 0"
Note="Organ: kidney; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;

-5' adaptor sequence: 5' GAATTCGGCAGG 3' ~3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Library provided
Lisa Guay-Woodford."
```

```

/db xref="taxon:10090"
/clone="F430014L12"
/tissue_type="spleen"
/dev_stage="6 days neonate"
/clone_lib="RIKEN full-length enriched, 6 days neonate spleen"

ORIGIN

Query Match      93.6%; Score 23.4; DB 2; Length 382;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCTCGG 25
    |||||
Db 67 AAAAAACGACACTGCGCGCCCTCGG 91

RESULT 39
BY629510
LOCUS
DEFINITION
cDNA clone K430031D24 3', mRNA sequence.
ACCESSION
BY629510
VERSION
EST.
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Santelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Haehizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

TITLE
Nature 420, 563-573 (2002)

JOURNAL
PUBLISHED
12466851
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

```

```

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN,
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
Location/Qualifiers
1..382
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430031D24"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match      93.6%; Score 23.4; DB 5; Length 382;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCTCGG 25
    |||||
Db 62 AAAAAACGACACTGCGCGCCCTCGG 86

RESULT 40
BY685406
LOCUS
DEFINITION
cDNA clone K430011M17 3', mRNA sequence.
ACCESSION
BY685406
VERSION
EST.
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Santelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Haehizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

TITLE
Nature 420, 563-573 (2002)

JOURNAL
PUBLISHED
12466851
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

```

Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, J. J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L. G., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

**TITLE**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL**  
Nature 420, 563-573 (2002)

**PUBMED**  
12466851

**COMMENT**  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Takashi Ishikawa (Department of Surgery 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan) whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

**FEATURES**  
source  
1. .382  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I320011M17"  
/cell\_type="stroma cell"  
/clone\_lib="RIKEN full-length enriched, stroma cell"

**ORIGIN**  
Query Match 93.6%; Score 23.4; DB 5; Length 382;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
Db 67 AAAAAACGACACGAGCCCTCGG 91  
|||||

**RESULT 41**

**BF235341**  
**LOCUS**  
**DEFINITION**  
602026734F1 NCI\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:4161752 5', mRNA sequence.  
**ACCESSION**  
BF235341  
**VERSION**  
BF235341.1  
**KEYWORDS**  
EST.  
**SOURCE**  
Mus musculus (house mouse)  
**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 383)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9443 row: d column: 09  
High quality sequence stop: 383.  
Location/Qualifiers  
1. .383  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4161752"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP Li9"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

**FEATURES**  
source  
1. .383  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4161752"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP Li9"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

**ORIGIN**  
Query Match 93.6%; Score 23.4; DB 2; Length 383;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCCCTCGG 25  
|||||  
Db 85 AAAAAACGACACGAGCCCTCGG 109  
|||||

**RESULT 42**  
BE197726/c  
**LOCUS**  
**DEFINITION**  
us77c09.x1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:3257200 3', mRNA sequence.  
**ACCESSION**  
BE197726  
**VERSION**  
BE197726.1  
**KEYWORDS**  
EST.  
**SOURCE**  
Mus musculus (house mouse)  
**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 384)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
image.lnl.gov/image/html/iresources.shtml

MGI:1068564

Seq primer: -40UP from Gibco  
High quality sequence stop: 383.

# FEATURES

Location/Qualifiers  
1. .384  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3257200"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH108"  
/clone\_lib="NCI CGAP Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Salt;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

# ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 384;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGCGAGCCCTCGG 25  
|||||  
Db 314 AAAAAACGACACTGCGAGCCCTCGG 290

# RESULT 43

BY629024 384 bp mRNA linear EST 15-DEC-2002  
LOCUS BY629024 RIKEN full-length enriched, visual cortex Mus musculus  
DEFINITION cDNA clone K430028L21 3', mRNA sequence.  
ACCESSION BY629024  
VERSION BY629024.1 GI:26964206  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

# ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 384)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saio, R., Suzuki, H., Yamana, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedziarski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Varardo, R., Wagner, L., Walstead, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,

Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

# JOURNAL

12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Pagliolini and Takao K. Hensch (  
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hiroawa, Wako-shi, Saitama 351-0198 Japan) whose  
assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

# FEATURES

Location/Qualifiers  
1. .384  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K430028L21"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

# ORIGIN

Query Match 93.6%; Score 23.4; DB 5; Length 384;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGCGAGCCCTCGG 25  
|||||  
Db 68 AAAAAACGACACTGCGAGCCCTCGG 92

# RESULT 44

BY699498 384 bp mRNA linear EST 16-DEC-2002  
LOCUS BY699498 RIKEN full-length enriched, osteoclast-like cell Mus  
DEFINITION musculus cDNA clone 1420117G23 3', mRNA sequence.  
ACCESSION BY699498  
VERSION BY699498.1 GI:27110532



KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus. 1 (bases 1 to 384)	
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL	Nature 420, 563-573 (2002)	
FURNED	12466851	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome.res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Takashi Ishikawa ( Department of Surgery 236-0004 Japan ) whose assistance we gratefully acknowledge. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.	
FEATURES	Location/Qualifiers 1..384 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="taxon:10090" /clone="I420117G23" /cell_type="osteoclast-like cell" /clone_lib="RIKEN full-length enriched, osteoclast-like cell"	Query Match 93.6%; Score 23.4; DB 5; Length 384; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
ORIGIN	QY 1 AAAAAACGACACGCGAGCCCTCGG 25 Db 69 AAAAAACGACACTCGCGAGCCCTCGG 93	
RESULT 45	BF320359/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	386 bp mRNA linear EST 29-DEC-2000 uz52h08.xl NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3672735 3', mRNA sequence. BF320359 BF320359 BF320359.1 GI:11269293 EST. Mus musculus (house mouse) Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus. 1 (bases 1 to 386) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Other ESTs: uz52h08.y1 Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey Green M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: image.llnl.gov/image/html/iresources.shtml MGI:1433503 High quality sequence stop: 385. Location/Qualifiers 1..386 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3672735" /sex="female, virgin" /tissue_type="infiltrating ductal carcinoma" /dev_stage="5 months" /clone_lib="NCI_CGAP Mam6" /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
FEATURES	Location/Qualifiers 1..386 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3672735" /sex="female, virgin" /tissue_type="infiltrating ductal carcinoma" /dev_stage="5 months" /clone_lib="NCI_CGAP Mam6" /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"	Query Match 93.6%; Score 23.4; DB 2; Length 386; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
ORIGIN	QY 93.6%; Score 23.4; DB 2; Length 386; Best Local Similarity 96.0%; Pred. No. 5.8; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	



Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michela Fagiolini and Takao K. Hensch (  
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose  
assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

FEATURES

Location/Qualifiers

1. .386

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="K430348K09"

/tissue\_type="visual cortex"

/clone\_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 93.6%; Score 23.4; DB 5; Length 386;

Best Local Similarity 96.0%; Pred. No. 5.8;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCGG 25

Db 78 AAAAAACGACACGAGCCCTCGG 102

RESULT 47

BY642036

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 AAAAAACGACACGAGCCCTCGG 25

11 AAAAAACGACACGAGCCCTCGG 31

12 AAAAAACGACACGAGCCCTCGG 42

13 AAAAAACGACACGAGCCCTCGG 53

14 AAAAAACGACACGAGCCCTCGG 64

15 AAAAAACGACACGAGCCCTCGG 75

16 AAAAAACGACACGAGCCCTCGG 86

17 AAAAAACGACACGAGCCCTCGG 97

18 AAAAAACGACACGAGCCCTCGG 108

19 AAAAAACGACACGAGCCCTCGG 119

20 AAAAAACGACACGAGCCCTCGG 130

21 AAAAAACGACACGAGCCCTCGG 141

22 AAAAAACGACACGAGCCCTCGG 152

23 AAAAAACGACACGAGCCCTCGG 163

24 AAAAAACGACACGAGCCCTCGG 174

25 AAAAAACGACACGAGCCCTCGG 185

26 AAAAAACGACACGAGCCCTCGG 196

27 AAAAAACGACACGAGCCCTCGG 207

28 AAAAAACGACACGAGCCCTCGG 218

29 AAAAAACGACACGAGCCCTCGG 229

30 AAAAAACGACACGAGCCCTCGG 240

31 AAAAAACGACACGAGCCCTCGG 251

32 AAAAAACGACACGAGCCCTCGG 262

33 AAAAAACGACACGAGCCCTCGG 273

34 AAAAAACGACACGAGCCCTCGG 284

35 AAAAAACGACACGAGCCCTCGG 295

36 AAAAAACGACACGAGCCCTCGG 306

37 AAAAAACGACACGAGCCCTCGG 317

38 AAAAAACGACACGAGCCCTCGG 328

39 AAAAAACGACACGAGCCCTCGG 339

40 AAAAAACGACACGAGCCCTCGG 350

41 AAAAAACGACACGAGCCCTCGG 361

42 AAAAAACGACACGAGCCCTCGG 372

43 AAAAAACGACACGAGCCCTCGG 383

44 AAAAAACGACACGAGCCCTCGG 394

45 AAAAAACGACACGAGCCCTCGG 405

46 AAAAAACGACACGAGCCCTCGG 416

47 AAAAAACGACACGAGCCCTCGG 427

48 AAAAAACGACACGAGCCCTCGG 438

49 AAAAAACGACACGAGCCCTCGG 449

50 AAAAAACGACACGAGCCCTCGG 460

51 AAAAAACGACACGAGCCCTCGG 471

52 AAAAAACGACACGAGCCCTCGG 482

53 AAAAAACGACACGAGCCCTCGG 493

54 AAAAAACGACACGAGCCCTCGG 504

55 AAAAAACGACACGAGCCCTCGG 515

56 AAAAAACGACACGAGCCCTCGG 526

57 AAAAAACGACACGAGCCCTCGG 537

58 AAAAAACGACACGAGCCCTCGG 548

59 AAAAAACGACACGAGCCCTCGG 559

60 AAAAAACGACACGAGCCCTCGG 570

61 AAAAAACGACACGAGCCCTCGG 581

62 AAAAAACGACACGAGCCCTCGG 592

63 AAAAAACGACACGAGCCCTCGG 603

64 AAAAAACGACACGAGCCCTCGG 614

65 AAAAAACGACACGAGCCCTCGG 625

66 AAAAAACGACACGAGCCCTCGG 636

67 AAAAAACGACACGAGCCCTCGG 647

68 AAAAAACGACACGAGCCCTCGG 658

69 AAAAAACGACACGAGCCCTCGG 669

70 AAAAAACGACACGAGCCCTCGG 680

71 AAAAAACGACACGAGCCCTCGG 691

72 AAAAAACGACACGAGCCCTCGG 702

73 AAAAAACGACACGAGCCCTCGG 713

74 AAAAAACGACACGAGCCCTCGG 724

75 AAAAAACGACACGAGCCCTCGG 735

76 AAAAAACGACACGAGCCCTCGG 746

77 AAAAAACGACACGAGCCCTCGG 757

78 AAAAAACGACACGAGCCCTCGG 768

79 AAAAAACGACACGAGCCCTCGG 779

80 AAAAAACGACACGAGCCCTCGG 790

81 AAAAAACGACACGAGCCCTCGG 801

82 AAAAAACGACACGAGCCCTCGG 812

83 AAAAAACGACACGAGCCCTCGG 823

84 AAAAAACGACACGAGCCCTCGG 834

85 AAAAAACGACACGAGCCCTCGG 845

86 AAAAAACGACACGAGCCCTCGG 856

87 AAAAAACGACACGAGCCCTCGG 867

88 AAAAAACGACACGAGCCCTCGG 878

89 AAAAAACGACACGAGCCCTCGG 889

90 AAAAAACGACACGAGCCCTCGG 900

91 AAAAAACGACACGAGCCCTCGG 911

92 AAAAAACGACACGAGCCCTCGG 922

93 AAAAAACGACACGAGCCCTCGG 933

94 AAAAAACGACACGAGCCCTCGG 944

95 AAAAAACGACACGAGCCCTCGG 955

96 AAAAAACGACACGAGCCCTCGG 966

97 AAAAAACGACACGAGCCCTCGG 977

98 AAAAAACGACACGAGCCCTCGG 988

99 AAAAAACGACACGAGCCCTCGG 999

100 AAAAAACGACACGAGCCCTCGG 1010

101 AAAAAACGACACGAGCCCTCGG 1021

102 AAAAAACGACACGAGCCCTCGG 1032

103 AAAAAACGACACGAGCCCTCGG 1043

104 AAAAAACGACACGAGCCCTCGG 1054

105 AAAAAACGACACGAGCCCTCGG 1065

106 AAAAAACGACACGAGCCCTCGG 1076

107 AAAAAACGACACGAGCCCTCGG 1087

108 AAAAAACGACACGAGCCCTCGG 1098

109 AAAAAACGACACGAGCCCTCGG 1109

110 AAAAAACGACACGAGCCCTCGG 1120

111 AAAAAACGACACGAGCCCTCGG 1131

112 AAAAAACGACACGAGCCCTCGG 1142

113 AAAAAACGACACGAGCCCTCGG 1153

114 AAAAAACGACACGAGCCCTCGG 1164

115 AAAAAACGACACGAGCCCTCGG 1175

116 AAAAAACGACACGAGCCCTCGG 1186

117 AAAAAACGACACGAGCCCTCGG 1197

118 AAAAAACGACACGAGCCCTCGG 1208

119 AAAAAACGACACGAGCCCTCGG 1219

120 AAAAAACGACACGAGCCCTCGG 1230

121 AAAAAACGACACGAGCCCTCGG 1241

122 AAAAAACGACACGAGCCCTCGG 1252

123 AAAAAACGACACGAGCCCTCGG 1263

124 AAAAAACGACACGAGCCCTCGG 1274

125 AAAAAACGACACGAGCCCTCGG 1285

126 AAAAAACGACACGAGCCCTCGG 1296

127 AAAAAACGACACGAGCCCTCGG 1307

128 AAAAAACGACACGAGCCCTCGG 1318

129 AAAAAACGACACGAGCCCTCGG 1329

130 AAAAAACGACACGAGCCCTCGG 1340

131 AAAAAACGACACGAGCCCTCGG 1351

132 AAAAAACGACACGAGCCCTCGG 1362

133 AAAAAACGACACGAGCCCTCGG 1373

134 AAAAAACGACACGAGCCCTCGG 1384

135 AAAAAACGACACGAGCCCTCGG 1395

136 AAAAAACGACACGAGCCCTCGG 1406

137 AAAAAACGACACGAGCCCTCGG 1417

138 AAAAAACGACACGAGCCCTCGG 1428

139 AAAAAACGACACGAGCCCTCGG 1439

140 AAAAAACGACACGAGCCCTCGG 1450

141 AAAAAACGACACGAGCCCTCGG 1461

142 AAAAAACGACACGAGCCCTCGG 1472

143 AAAAAACGACACGAGCCCTCGG 1483

144 AAAAAACGACACGAGCCCTCGG 1494

145 AAAAAACGACACGAGCCCTCGG 1505

146 AAAAAACGACACGAGCCCTCGG 1516

147 AAAAAACGACACGAGCCCTCGG 1527

148 AAAAAACGACACGAGCCCTCGG 1538

149 AAAAAACGACACGAGCCCTCGG 1549

150 AAAAAACGACACGAGCCCTCGG 1560

151 AAAAAACGACACGAGCCCTCGG 1571

152 AAAAAACGACACGAGCCCTCGG 1582

153 AAAAAACGACACGAGCCCTCGG 1593

154 AAAAAACGACACGAGCCCTCGG 1604

155 AAAAAACGACACGAGCCCTCGG 1615

156 AAAAAACGACACGAGCCCTCGG 1626

157 AAAAAACGACACGAGCCCTCGG 1637

158 AAAAAACGACACGAGCCCTCGG 1648

159 AAAAAACGACACGAGCCCTCGG 1659

160 AAAAAACGACACGAGCCCTCGG 1670

161 AAAAAACGACACGAGCCCTCGG 1681

162 AAAAAACGACACGAGCCCTCGG 1692

163 AAAAAACGACACGAGCCCTCGG 1703

164 AAAAAACGACACGAGCCCTCGG 1714

165 AAAAAACGACACGAGCCCTCGG 1725

166 AAAAAACGACACGAGCCCTCGG 1736

167 AAAAAACGACACGAGCCCTCGG 1747

168 AAAAAACGACACGAGCCCTCGG 1758

169 AAAAAACGACACGAGCCCTCGG 1769

170 AAAAAACGACACGAGCCCTCGG 1780

171 AAAAAACGACACGAGCCCTCGG 1791

172 AAAAAACGACACGAGCCCTCGG 1802

173 AAAAAACGACACGAGCCCTCGG 1813

174 AAAAAACGACACGAGCCCTCGG 1824

175 AAAAAACGACACGAGCCCTCGG 1835

176 AAAAAACGACACGAGCCCTCGG 1846

177 AAAAAACGACACGAGCCCTCGG 1857

178 AAAAAACGACACGAGCCCTCGG 1868

179 AAAAAACGACACGAGCCCTCGG 1879

180 AAAAAACGACACGAGCCCTCGG 1890

181 AAAAAACGACACGAGCCCTCGG 1901

182 AAAAAACGACACGAGCCCTCGG 1912

183 AAAAAACGACACGAGCCCTCGG 1923

184 AAAAAACGACACGAGCCCTCGG 1934

185 AAAAAACGACACGAGCCCTCGG 1945

186 AAAAAACGACACGAGCCCTCGG 1956

187 AAAAAACGACACGAGCCCTCGG 1967

188 AAAAAACGACACGAGCCCTCGG 1978

189 AAAAAACGACACGAGCCCTCGG 1989

190 AAAAAACGACACGAGCCCTCGG 2000

191 AAAAAACGACACGAGCCCTCGG 2011

192 AAAAAACGACACGAGCCCTCGG 2022

193 AAAAAACGACACGAGCCCTCGG 2033

194 AAAAAACGACACGAGCCCTCGG 2044

195 AAAAAACGACACGAGCCCTCGG 2055

196 AAAAAACGACACGAGCCCTCGG 2066

197 AAAAAACGACACGAGCCCTCGG 2077

198 AAAAAACGACACGAGCCCTCGG 2088

199 AAAAAACGACACGAGCCCTCGG 2099

200 AAAAAACGACACGAGCCCTCGG 2110

201 AAAAAACGACACGAGCCCTCGG 2121

202 AAAAAACGACACGAGCCCTCGG 2132

203 AAAAAACGACACGAGCCCTCGG 2143

204 AAAAAACGACACGAGCCCTCGG 2154

205 AAAAAACGACACGAGCCCTCGG 2165

206 AAAAAACGACACGAGCCCTCGG 2176

207 AAAAAACGACACGAGCCCTCGG 2187

208 AAAAAACGACACGAGCCCTCGG 2198

209 AAAAAACGACACGAGCCCTCGG 2209

210 AAAAAACGACACGAGCCCTCGG 2220

211 AAAAAACGACACGAGCCCTCGG 2231

212 AAAAAACGACACGAGCCCTCGG 2242

213 AAAAAACGACACGAGCCCTCGG 2253

214 AAAAAACGACACGAGCCCTCGG 2264

215 AAAAAACGACACGAGCCCTCGG 2275

216 AAAAAACGACACGAGCCCTCGG 2286

217 AAAAAACGACACGAGCCCTCGG 2297

218 AAAAAACGACACGAGCCCTCGG 2308

219 AAAAAACGACACGAGCCCTCGG 2319

220 AAAAAACGACACGAGCCCTCGG 2330

221 AAAAAACGACACGAGCCCTCGG 2341

222 AAAAAACGACACGAGCCCTCGG 2352

223 AAAAAACGACACGAGCCCTCGG 2363

224 AAAAAACGACACGAGCCCTCGG 2374

225 AAAAAACGACACGAGCCCTCGG 2385

226 AAAAAACGACACGAGCCCTCGG 2396

227 AAAAAACGACACGAGCCCTCGG 2407

228 AAAAAACGACACGAGCCCTCGG 2418

229 AAAAAACGACACGAGCCCTCGG 2429

230 AAAAAACGACACGAGCCCTCGG 2440

231 AAAAAACGACACGAGCCCTCGG 2451

232 AAAAAACGACACGAGCCCTCGG 2462

233 AAAAAACGACACGAGCCCTCGG 2473

234 AAAAAACGACACGAGCCCTCGG 2484

235 AAAAAACGACACGAGCCCTCGG 2495

236 AAAAAACGACACGAGCCCTCGG 2506

237 AAAAAACGACACGAGCCCTCGG 2517

238 AAAAAACGACACGAGCCCTCGG 2528

239 AAAAAACGACACGAGCCCTCGG 2539

240 AAAAAACGACACGAGCCCTCGG 2550

241 AAAAAACGACACGAGCCCTCGG 2561

242 AAAAAACGACACGAGCCCTCGG 2572

243 AAAAAACGACACGAGCCCTCGG 2583

244 AAAAAACGACACGAGCCCTCGG 2594

245 AAAAAACGACACGAGCCCTCGG 2605

246 AAAAAACGACACGAGCCCTCGG 2616

247 AAAAAACGACACGAGCCCTCGG 2627

248 AAAAAACGACACGAGCCCTCGG 2638

249 AAAAAACGACACGAGCCCTCGG 2649

250 AAAAAACGACACGAGCCCTCGG 2660

251 AAAAAACGACACGAGCCCTCGG 2671

252 AAAAAACGACACGAGCCCTCGG 2682

253 AAAAAACGACACGAGCCCTCGG 2693

254 AAAAAACGACACGAGCCCTCGG 2704

255 AAAAAACGACACGAGCCCTCGG 2715

256 AAAAAACGACACGAGCCCTCGG 2726

257 AAAAAACGACACGAGCCCTCGG 2737

258 AAAAAACGACACGAGCCCTCGG 2748

259 AAAAAACGACACGAGCCCTCGG 2759

260 AAAAAACGACACGAGCCCTCGG 2770

261 AAAAAACGACACGAGCCCTCGG 2781

262 AAAAAACGACACGAGCCCTCGG 2792

263 AAAAAACGACACGAGCCCTCGG 2803

264 AAAAAACGACACGAGCCCTCGG 2814

265 AAAAAACGACACGAGCCCTCGG 2825

266 AAAAAACGACACGAGCCCTCGG 2836

267 AAAAAACGACACGAGCCCTCGG 2847

268 AAAAAACGACACGAGCCCTCGG 2858

269 AAAAAACGACACGAGCCCTCGG 2869

270 AAAAAACGACACGAGCCCTCGG 2880

271 AAAAAACGACACGAGCCCTCGG 2891

272 AAAAAACGACACGAGCCCTCGG 2902

273 AAAAAACGACACGAGCCCTCGG 2913

274 AAAAAACGACACGAGCCCTCGG 2924

275 AAAAAACGACACGAGCCCTCGG 2935

276 AAAAAACGACACGAGCCCTCGG 2946

277 AAAAAACGACACGAGCCCTCGG 2957

278 AAAAAACGACACGAGCCCTCGG 2968

279 AAAAAACGACACGAGCCCTCGG 2979

280 AAAAAACGACACGAGCCCTCGG 2990

281 AAAAAACGACACGAGCCCTCGG 3001

282 AAAAAACGACACGAGCCCTCGG 3012

283 AAAAAACGACACGAGCCCTCGG 3023

284 AAAAAACGACACGAGCCCTCGG 3034

285 AAAAAACGACACGAGCCCTCGG 3045

286 AAAAAACGACACGAGCCCTCGG 3056

287 AAAAAACGACACGAGCCCTCGG 3067

288 AAAAAACGACACGAGCCCTCGG 3078

289 AAAAAACGACACGAGCCCTCGG 3089

290 AAAAAACGACACGAGCCCTCGG 3100

291 AAAAAACGACACGAGCCCTCGG 3111

292 AAAAAACGACACGAGCCCTCGG 3122

293 AAAAAACGACACGAGCCCTCGG 3133

294 AAAAAACGACACGAGCCCTCGG 3144

295 AAAAAACGACACGAGCCCTCGG 3155

296 AAAAAACGACACGAGCCCTCGG 3166

297 AAAAAACGACACGAGCCCTCGG 3177

298 AAAAAACGACACGAGCCCTCGG 3188

299 AAAAAACGACACGAGCCCTCGG 3199

300 AAAAAACGACACGAGCCCTCGG 3210

301 AAAAAACGACACGAGCCCTCGG 3221

302 AAAAAACGACACGAGCCCTCGG 3232

303 AAAAAACGACACGAGCCCTCGG 3243

304 AAAAAACGACACGAGCCCTCGG 3254

305 AAAAAACGACACGAGCCCTCGG 3265

306 AAAAAACGACACGAGCCCTCGG 3276

307 AAAAAACGACACGAGCCCTCGG 3287

308 AAAAAACGACACGAGCCCTCGG 3298

309 AAAAAACGACACGAGCCCTCGG 3309

310 AAAAAACGACACGAGCCCTCGG 3320

311 AAAAAACGACACGAGCCCTCGG 3331

312 AAAAAACGACACGAGCCCTCGG 3342

313 AAAAAACGACACGAGCCCTCGG 3353

314 AAAAAACGACACGAGCCCTCGG 3364

315 AAAAAACGACACGAGCCCTCGG 3375

316 AAAAAACGACACGAGCCCTCGG 3386

317 AAAAAACGACACGAGCCCTCGG 3397

318 AAAAAACGACACGAGCCCTCGG 3408

319 AAAAAACGACACGAGCCCTCGG 3419

320 AAAAAACGACACGAGCCCTCGG 3430

321 AAAAAACGACACGAGCCCTCGG 3441

322 AAAAAACGACACGAGCCCTCGG 3452

323 AAAAAACGACACGAGCCCTCGG 3463

324 AAAAAACGACACGAGCCCTCGG 3474

325 AAAAAACGACACGAGCCCTCGG 3485

326 AAAAAACGACACGAGCCCTCGG 3496

327 AAAAAACGACACGAGCCCTCGG 3507

328 AAAAAACGACACGAGCCCTCGG 3518

329 AAAAAACGACACGAGCCCTCGG 3529

330 AAAAAACGACACGAGCCCTCGG 3540

331 AAAAAACGACACGAGCCCTCGG 3551

332 AAAAAACGACACGAGCCCTCGG 3562

333 AAAAAACGACACGAGCCCTCGG 3573

334 AAAAAACGACACGAGCCCTCGG 3584

335 AAAAAACGACACGAGCCCTCGG 3595

336 AAAAAACGACACGAGCCCTCGG 3606

337 AAAAAACGACACGAGCCCTCGG 3617

338 AAAAAACGACACGAGCCCTCGG 3628

339 AAAAAACGACACGAGCCCTCGG 3639

340 AAAAAACGACACGAGCCCTCGG 3650

341 AAAAAACGACACGAGCCCTCGG 3661

342 AAAAAACGACACGAGCCCTCGG 3672

343 AAAAAACGACACGAGCCCTCGG 3683

344 AAAAAACGACACGAGCCCTCGG 3694

345 AAAAAACGACACGAGCCCTCGG 3705

346 AAAAAACGACACGAGCCCTCGG 3716

347 AAAAAACGACACGAGCCCTCGG 3727

348 AAAAAACGACACGAGCCCTCGG 3738

349 AAAAAACGACACGAGCCCTCGG 3749

350 AAAAAACGACACGAGCCCTCGG 3760

351 AAAAAACGACACGAGCCCTCGG 3771

352 AAAAAACGACACGAGCCCTCGG 3782

353 AAAAAACGACACGAGCCCTCGG 3793

354 AAAAAACGACACGAGCCCTCGG 3804

355 AAAAAACGACACGAGCCCTCGG 3815

356 AAAAAACGACACGAGCCCTCGG 3826

357 AAAAAACGACACGAGCCCTCGG 3837

358 AAAAAACGACACGAGCCCTCGG 3848

359 AAAAAACGACACGAGCCCTCGG 3859

360 AAAAAACGACACGAGCCCTCGG 3870

361 AAAAAACGACACGAGCCCTCGG 3881

362 AAAAAACGACACGAGCCCTCGG 3892

363 AAAAAACGACACGAGCCCTCGG 3903

364 AAAAAACGACACGAGCCCTCGG 3914

365 AAAAAACGACACGAGCCCTCGG 3925

366 AAAAAACGACACGAGCCCTCGG 3936

367 AAAAAACGACACGAGCCCTCGG 3947

368 AAAAAACGACACGAGCCCTCGG 3958

369 AAAAAACGACACGAGCCCTCGG 3969

370 AAAAAACGACACGAGCCCTCGG 3980

371 AAAAAACGACACGAGCCCTCGG 3991

372 AAAAAACGACACGAGCCCTCGG 4002

373 AAAAAACGACACGAGCCCTCGG 4013

374 AAAAAACGACACGAGCCCTCGG 4024

375 AAAAAACGACACGAGCCCTCGG 4035

376 AAAAAACGACACGAGCCCTCGG 4046

377 AAAAAACGACACGAGCCCTCGG 4057

378 AAAAAACGACACGAGCCCTCGG 4068

379 AAAAAACGACACGAGCCCTCGG 4079

380 AAAAAACGACACGAGCCCTCGG 4090

381 AAAAAACGACACGAGCCCTCGG 4101

382 AAAAAACGACACGAGCCCTCGG 4112

383 AAAAAACGACACGAGCCCTCGG 4123

384 AAAAAACGACACGAGCCCTCGG 4134

385 AAAAAACGACACGAGCCCTCGG 4145

386 AAAAAACGACACGAGCCCTCGG 4156

387 AAAAAACGACACGAGCCCTCGG 4167

388 AAAAAACGACACGAGCCCTCGG 4178

389 AAAAAACGACACGAGCCCTCGG 4189

390 AAAAAACGACACGAGCCCTCGG 4200

391 AAAAAACGACACGAGCCCTCGG 4211

392 AAAAAACGACACGAGCCCTCGG 4222

393 AAAAAACGACACGAGCCCTCGG 4233

394 AAAAAACGACACGAGCCCTCGG 4244

395 AAAAAACGACACGAGCCCTCGG 4255

396 AAAAAACGACACGAGCCCTCGG 4266

397 AAAAAACGACACGAGCCCTCGG 4277

398 AAAAAACGACACGAGCCCTCGG 4288

399 AAAAAACGACACGAGCCCTCGG 4299

400 AAAAAACGACACGAGCCCTCGG 4310

401 AAAAAACGACACGAGCCCTCGG 4321

402 AAAAAACGACACGAGCCCTCGG 4332

403 AAAAAACGACACGAGCCCTCGG 4343

404 AAAAAACGACACGAGCCCTCGG 4354

405 AAAAAACGACACGAGCCCTCGG 4365

406 AAAAAACGACACGAGCCCTCGG 4376

407 AAAAAACGACACGAGCCCTCGG 4387

408 AAAAAACGACACGAGCCCTCGG 4398

409 AAAAAACGACACGAGCCCTCGG 4409

410 AAAAAACGACACGAGCCCTCGG 4420

411 AAAAAACGACACGAGCCCTCGG 4431

412 AAAAAACGACACGAGCCCTCGG 4442

413 AAAAAACGACACGAGCCCTCGG 4453

414 AAAAAACGACACGAGCCCTCGG 4464

415 AAAAAACGACACGAGCCCTCGG 4475

416 AAAAAACGACACGAGCCCTCGG 4486

417 AAAAAACGACACGAGCCCTCGG 4497

418 AAAAAACGACACGAGCCCTCGG 4508

419 AAAAAACGACACGAGCCCTCGG 4519

420 AAAAAACGACACGAGCCCTCGG 4530

421 AAAAAACGACACGAGCCCTCGG 4541

422 AAAAAACGACACGAGCCCTCGG 4552

423 AAAAAACGACACGAGCCCTCGG 4563

424 AAAAAACGACACGAGCCCTCGG 4574

425 AAAAAACGACACGAGCCCTCGG 4585

426 AAAAAACGACACGAGCCCTCGG 4596

427 AAAAAACGACACGAGCCCTCGG 4607

428 AAAAAACGACACGAGCCCTCGG 4618

429 AAAAAACGACACGAGCCCTCGG 4629

430 AAAAAACGACACGAGCCCTCGG 4640

431 AAAAAACGACACGAGCCCTCGG 4651

432 AAAAAACGACACGAGCCCTCGG 4662

433 AAAAAACGACACGAGCCCTCGG 4673

434 AAAAAACGACACGAGCCCTCGG 4684

435 AAAAAACGACACGAGCCCTCGG 4695

436 AAAAAACGACACGAGCCCTCGG 4706

437 AAAAAACGACACGAGCCCTCGG 4717

438 AAAAAACGACACGAGCCCTCGG 4728

439 AAAAAACGACACGAGCCCTCGG 4739

440 AAAAAACGACACGAGCCCTCGG 4750

441 AAAAAACGACACGAGCCCTCGG 4761

442 AAAAAACGACACGAGCCCTCGG 4772

443 AAAAAACGACACGAGCCCTCGG 4783

444 AAAAAACGACACGAGCCCTCGG 4794

445 AAAAAACGACACGAGCCCTCGG 4805

446 AAAAAACGACACGAGCCCTCGG 4816

447 AAAAAACGACACGAGCCCTCGG 4827

448 AAAAAACGACACGAGCCCTCGG 4838

449 AAAAAACGACACGAGCCCTCGG 4849

450 AAAAAACGACACGAGCCCTCGG 4860

451 AAAAAACGACACGAGCCCTCGG 4871

452 AAAAAACGACACGAGCCCTCGG 4882

453 AAAAAACGACACGAGCCCTCGG 4893

454 AAAAAACGACACGAGCCCTCGG 4904

455 AAAAAACGACACGAGCCCTCGG 4915

456 AAAAAACGACACGAGCCCTCGG 4926

457 AAAAAACGACACGAGCCCTCGG 4937

458 AAAAAACGACACGAGCCCTCGG 4948

459 AAAAAACGACACGAGCCCTCGG 4959

460 AAAAAACGACACGAGCCCTCGG 4970

461 AAAAAACGACACGAGCCCTCGG 4981

462 AAAAAACGACACGAGCCCTCGG 4992

463 AAAAAACGACACGAGCCCTCGG 5003

464 AAAAAACGACACGAGCCCTCGG 5014

465 AAAAAACGACACGAGCCCTCGG 5025

466 AAAAAACGACACGAGCCCTCGG 5036

467 AAAAAACGACACGAGCCCTCGG 5047

468 AAAAAACGACACGAGCCCTCGG 5058

469 AAAAAACGACACGAGCCCTCGG 5069

470 AAAAAACGACACGAGCCCTCGG 5080

471 AAAAAACGACACGAGCCCTCGG 5091

472 AAAAAACGACACGAGCCCTCGG 5102

473 AAAAAACGACACGAGCCCTCGG 5113

474 AAAAAACGACACGAGCCCTCGG 5124

475 AAAAAACGACACGAGCCCTCGG 5135

476 AAAAAACGACACGAGCCCTCGG 5146

477 AAAAAACGACACGAGCCCTCGG 5157

478 AAAAAACGACACGAGCCCTCGG 5168

479 AAAAAACGACACGAGCCCTCGG 5179

480 AAAAAACGACACGAGCCCTCGG 5190

481 AAAAAACGACACGAGCCCTCGG 5201

482 AAAAAACGACACGAGCCCTCGG 5212

483 AAAAAACGACACGAGCCCTCGG 5223

484 AAAAAACGACACGAGCCCTCGG 5234

485 AAAAAACGACACGAGCCCT

Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851

## TITLE

JOURNAL  
PUBMED

## COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Fagioli and Takao K. Henach (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hiroawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

## FEATURES

source

Location/Qualifiers  
1. .387  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K230028E20"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

## ORIGIN

Query Match . 93.6%; Score 23.4; DB 5; Length 387;  
Best Local Similarity 96.0%; Pred. No. 5.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACAGCGAGCCTCGG 25

|||||

71 AAAAAACGACAGCGAGCCTCGG 95

## RESULT 48

BY630813

## LOCUS

BY630813 RIKEN full-length enriched, visual cortex Mus musculus  
cDNA clone K430038M12 3', mRNA sequence.

## ACCESSION

BY630813

## VERSION

BY630813.1

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

## AUTHORS

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 397)

Okazaki,I., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,  
Nikaido,I., Oeato,N., Saito,R., Suzuki,H., Yamataka,I.,  
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,  
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,  
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuoka,H.,  
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,  
Chochlov,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,  
Fletcher,C.F., Forrest,A., Frazer,K.S., Gasteland,I.,  
Garcinovich,S., Gliori,C., Godzik,A., Gough,J., Grimmond,S.,  
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,  
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,  
Kurochin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,  
Maltas,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,  
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,  
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramchandran,S.,  
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,  
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,  
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,  
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,  
Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yanagisawa,M., Yang,I.,  
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,  
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,  
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,  
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851

## TITLE

## JOURNAL

## PUBMED

## COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Fagioli and Takao K. Henach (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hiroawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

## FEATURES

Location/Qualifiers

1. .387

/organism="Mus musculus"

```

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430038M12"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match      93.6%; Score 23.4; DB 5; Length 387;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25
    |||||
Db 74 AAAAAACGACACGAGCGCCCTCGG 98

RESULT 49
BY660055
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
  Nishikado, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I.,
  Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
  Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
  Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
  Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
  Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
  Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
  Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
  Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
  Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
  Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
  Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
  Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
  Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
  Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
  Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
  Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
  Varardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
  Wells, C., Wilming, L.G., Wyszynski, B., Yang, I., Yang, I.,
  Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
  Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
  Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
  Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
  Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
  Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
  Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
PUBMED
COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watanishi, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagioli and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirozawa, Wako-shi, Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
Location/Qualifiers
1..387
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K530360P19"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match      93.6%; Score 23.4; DB 5; Length 387;
Best Local Similarity 96.0%; Pred. No. 5.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCCCTCGG 25
    |||||
Db 72 AAAAAACGACACGAGCGCCCTCGG 96

RESULT 50
BY660055
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
  Nishikado, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I.,
  Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
  Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
  Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
  Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
  Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
  Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
  Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
  Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
  Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
  Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
  Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
  Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
  Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
  Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
  Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
  Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
  Varardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
  Wells, C., Wilming, L.G., Wyszynski, B., Yang, I., Yang, I.,
  Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
  Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
  Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
  Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
  Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
  Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
  Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
PUBMED
COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watanishi, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagioli and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirozawa, Wako-shi, Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

```

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-15

Perfect score: 25

Sequence: 1 aaaaacacacagcagccctcg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.6	70.4	32881	3	US-09-949-016-16048
C 2	17	68.0	601	3	US-09-949-016-59715
C 3	17	68.0	603	3	US-09-489-039A-6163
C 4	17	68.0	19857	3	US-09-949-016-13114
C 5	17	68.0	19998	3	US-09-949-016-12657
C 6	17	68.0	36938	3	US-09-949-016-13484
C 7	16.6	66.4	255	3	US-09-248-796A-11154
C 8	16.6	66.4	601	3	US-09-949-002-1163
C 9	16.6	66.4	601	3	US-09-949-002-7259
C 10	16.6	66.4	2025	3	US-08-991-677-7
C 11	16.6	66.4	4493	3	US-09-949-016-14034
C 12	16.6	66.4	40325	3	US-09-949-002-580
C 13	16.6	66.4	40325	3	US-09-949-002-771
C 14	16.6	66.4	43717	3	US-09-949-016-16821
C 15	16.2	64.8	452	3	US-09-300-958A-40
C 16	16.2	64.8	1409	9	5242798-6
C 17	16.2	64.8	2826	3	US-09-902-540-3646
C 18	16.2	64.8	16563	3	US-09-902-540-2000
C 19	16.2	64.8	16844	3	US-09-902-540-1176
C 20	16.2	64.8	28493	3	US-09-902-540-1241
C 21	16.2	64.8	41322	3	US-10-024-396-13
C 22	16.2	64.8	90472	3	US-09-949-016-14038
C 23	16.2	64.8	158735	3	US-09-949-016-11989
C 24	16.2	64.8	158735	3	US-09-949-016-17130

C 25	16	64.0	184	3	US-09-513-999C-16876	Sequence 16876, A
C 26	16	64.0	531	3	US-09-533-559-2241	Sequence 2241, Ap
C 27	16	64.0	601	3	US-09-949-016-27169	Sequence 27169, A
C 28	16	64.0	601	3	US-09-949-016-178692	Sequence 178692, A
C 29	16	64.0	601	3	US-09-949-016-178693	Sequence 178693, A
C 30	16	64.0	601	3	US-09-949-016-182367	Sequence 182367, A
C 31	16	64.0	1347	3	US-09-543-681A-2689	Sequence 2689, Ap
C 32	16	64.0	1600	3	US-09-949-016-651	Sequence 651, App
C 33	16	64.0	13595	3	US-09-949-016-12529	Sequence 12529, A
C 34	16	64.0	13970	3	US-09-949-016-16690	Sequence 16690, A
C 35	16	64.0	27968	3	US-09-949-016-15191	Sequence 15191, A
C 36	16	64.0	27968	3	US-09-949-016-15192	Sequence 15192, A
C 37	16	64.0	34372	3	US-09-949-016-13098	Sequence 13098, A
C 38	16	64.0	34875	3	US-09-949-016-13099	Sequence 13099, A
C 39	16	64.0	37861	3	US-09-949-016-16872	Sequence 16872, A
C 40	16	64.0	97195	3	US-09-949-016-12212	Sequence 12212, A
C 41	16	64.0	97196	3	US-09-949-016-16971	Sequence 16971, A
C 42	15.8	63.2	571	3	US-08-779-764A-13	Sequence 13, Appl
C 43	15.8	63.2	571	3	US-09-563-456-13	Sequence 13, Appl
C 44	15.8	63.2	1692	3	US-09-252-991A-7872	Sequence 7872, Ap
C 45	15.8	63.2	6755	3	US-08-931-999-4	Sequence 4, Appl1
C 46	15.8	63.2	8910	3	US-08-779-764A-1	Sequence 1, Appl1
C 47	15.8	63.2	8910	3	US-09-563-456-1	Sequence 1, Appl1
C 48	15.8	63.2	72662	3	US-09-949-016-17073	Sequence 17073, A
C 49	15.8	63.2	131254	3	US-09-949-016-13734	Sequence 13734, A
C 50	15.8	63.2	135030	3	US-09-949-016-14896	Sequence 14896, A
C 51	15.6	62.4	111	3	US-09-702-705-1556	Sequence 1556, Ap
C 52	15.6	62.4	111	3	US-09-736-457-1556	Sequence 1556, Ap
C 53	15.6	62.4	111	3	US-09-614-1248-1556	Sequence 1556, Ap
C 54	15.6	62.4	111	3	US-09-671-325-1556	Sequence 1556, Ap
C 55	15.6	62.4	111	3	US-09-658-824-1556	Sequence 1556, Ap
C 56	15.6	62.4	111	3	US-10-017-754-1556	Sequence 1556, Ap
C 57	15.6	62.4	111	3	US-09-651-563-1556	Sequence 1556, Ap
C 58	15.6	62.4	258	3	US-09-543-681A-1487	Sequence 1487, Ap
C 59	15.6	62.4	309	3	US-09-513-999C-32302	Sequence 32302, A
C 60	15.6	62.4	315	3	US-09-621-976-17895	Sequence 17895, A
C 61	15.6	62.4	414	3	US-08-905-223-191	Sequence 191, App
C 62	15.6	62.4	426	3	US-09-621-976-17892	Sequence 17892, A
C 63	15.6	62.4	466	3	US-09-513-999C-886	Sequence 886, App
C 64	15.6	62.4	601	3	US-09-949-016-78047	Sequence 78047, A
C 65	15.6	62.4	601	3	US-09-949-016-139389	Sequence 139389, A
C 66	15.6	62.4	601	3	US-09-949-016-139390	Sequence 139390, A
C 67	15.6	62.4	601	3	US-09-949-016-139391	Sequence 139391, A
C 68	15.6	62.4	601	3	US-09-949-016-139392	Sequence 139392, A
C 69	15.6	62.4	601	3	US-09-949-016-169403	Sequence 169403, A
C 70	15.6	62.4	610	2	US-08-928-926A-2	Sequence 2, Appl1
C 71	15.6	62.4	610	3	US-09-212-149-2	Sequence 2, Appl1
C 72	15.6	62.4	654	3	US-09-543-681A-147	Sequence 147, App
C 73	15.6	62.4	954	3	US-09-252-991A-15575	Sequence 15575, A
C 74	15.6	62.4	1098	3	US-09-774-528-183	Sequence 183, App
C 75	15.6	62.4	1098	3	US-10-120-988-183	Sequence 183, App
C 76	15.6	62.4	1116	3	US-09-252-991A-15677	Sequence 15677, A
C 77	15.6	62.4	5999	3	US-09-949-016-13836	Sequence 13836, A
C 78	15.6	62.4	22294	3	US-09-949-016-14020	Sequence 14020, A
C 79	15.6	62.4	45432	3	US-09-949-002-719	Sequence 719, App
C 80	15.6	62.4	80595	3	US-09-078-294-4	Sequence 4, Appl1
C 81	15.6	62.4	80595	3	US-09-949-016-15684	Sequence 15684, A
C 82	15.6	62.4	117001	3	US-09-949-016-14982	Sequence 14982, A
C 83	15.6	62.4	127771	3	US-09-949-016-14982	Sequence 14982, A
C 84	15.6	62.4	17293	3	US-09-949-016-16313	Sequence 16313, A
C 85	15.6	62.4	258775	3	US-09-949-016-16435	Sequence 16435, A
C 86	15.4	61.6	232	3	US-09-513-999C-33162	Sequence 33162, A
C 87	15.4	61.6	375	3	US-09-270-767-15003	Sequence 15003, A
C 88	15.4	61.6	462	3	US-09-270-767-6285	Sequence 6285, Ap
C 89	15.4	61.6	462	3	US-09-270-767-21567	Sequence 21567, A
C 90	15.4	61.6	488	3	US-09-270-767-4242	Sequence 4242, Ap
C 91	15.4	61.6	488	3	US-09-270-767-19524	Sequence 19524, A
C 92	15.4	61.6	494	3	US-09-533-559-5232	Sequence 5232, Ap
C 93	15.4	61.6	601	3	US-09-949-016-42914	Sequence 42914, A
C 94	15.4	61.6	601	3	US-09-949-016-42915	Sequence 42915, A
C 95	15.4	61.6	601	3	US-09-949-016-42916	Sequence 42916, A
C 96	15.4	61.6	601	3	US-09-949-016-59081	Sequence 59081, A
C 97	15.4	61.6	601	3	US-09-949-016-59082	Sequence 59082, A

C 98 15.4 61.6 601 3 US-09-949-016-109227  
99 15.4 61.6 601 3 US-09-949-016-119333  
100 15.4 61.6 601 3 US-09-949-016-119334  
C 101 15.4 61.6 800 3 US-08-945-809A-39  
C 102 15.4 61.6 800 3 US-09-515-697-39  
103 15.4 61.6 842 3 US-10-144-929-36  
C 104 15.4 61.6 984 3 US-09-543-681A-2047  
105 15.4 61.6 1181 3 US-09-533-559-6100  
106 15.4 61.6 1437 3 US-09-540-236-711  
C 107 15.4 61.6 1590 3 US-09-489-039A-2550  
C 108 15.4 61.6 2364 3 US-09-489-039A-457  
109 15.4 61.6 10660 2 US-08-267-803B-8  
110 15.4 61.6 10660 2 US-09-041-886-16  
111 15.4 61.6 11517 2 US-07-920-281C-1  
112 15.4 61.6 11517 3 US-08-466-277-1  
113 15.4 61.6 11517 3 US-09-688-842-1  
C 114 15.4 61.6 14524 3 US-09-949-016-12552  
115 15.4 61.6 14785 3 US-09-949-016-15435  
C 116 15.4 61.6 15000 3 US-09-719-554-65  
117 15.4 61.6 20182 3 US-09-949-016-13240  
C 118 15.4 61.6 21410 3 US-09-596-002-12  
C 119 15.4 61.6 25190 3 US-09-949-016-15906  
C 120 15.4 61.6 34589 3 US-09-949-002-836  
C 121 15.4 61.6 37412 3 US-09-949-016-17566  
122 15.4 61.6 42394 3 US-09-949-016-12752  
123 15.4 61.6 42395 3 US-09-949-016-15115  
124 15.4 61.6 43950 3 US-09-735-934A-3  
125 15.4 61.6 43950 3 US-10-060-332-3  
126 15.4 61.6 43950 3 US-10-339-657-3  
127 15.4 61.6 43950 3 US-10-885-879-3  
C 128 15.4 61.6 49378 3 US-09-949-016-13408  
C 129 15.4 61.6 59258 3 US-09-949-002-581  
C 130 15.4 61.6 61777 3 US-09-949-016-17278  
C 131 15.4 61.6 83708 3 US-09-949-016-17207  
C 132 15.4 61.6 89716 3 US-09-949-016-11900  
C 133 15.4 61.6 106924 3 US-09-949-016-13934  
C 134 15.4 61.6 119801 3 US-09-949-016-13453  
C 135 15.4 61.6 135657 3 US-09-949-016-15051  
C 136 15.4 61.6 152486 3 US-09-949-016-12869  
137 15.4 61.6 154600 3 US-09-949-016-14757  
C 138 15.4 61.6 175265 3 US-09-949-016-16089  
C 139 15.4 61.6 179904 3 US-09-949-002-577  
C 140 15.4 61.6 179905 3 US-09-949-002-705  
141 15.4 61.6 187916 3 US-09-949-016-12980  
C 142 15.4 61.6 227979 3 US-09-949-016-11842  
143 15.4 61.6 236964 3 US-09-949-016-15753  
144 15.4 61.6 312957 3 US-09-949-001-31  
145 15.4 61.6 312972 3 US-09-949-001-34  
C 146 15.4 61.6 784019 3 US-09-949-016-14033  
C 147 15.4 61.6 828152 3 US-09-949-016-12777  
C 148 15.2 60.8 504 3 US-09-270-767-6856  
C 149 15.2 60.8 504 3 US-09-270-767-22138  
150 15.2 60.8 527 3 US-09-640-211A-1849

ALIGNMENTS

RESULT 1  
US-09-949-016-16048/c  
; Sequence 16048, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

RESULT 3  
US-09-949-039A-6163/c  
; Sequence 6163, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 6163  
; LENGTH: 603

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 476 AAGATCCACACAGTGAGCCCTCTG 500

RESULT 2  
US-09-949-016-59715  
; Sequence 59715, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59715  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-59715

Query Match 70.4%; Score 17.6; DB 3; Length 32881;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAPACGACACAGCGAGCCCTCGG 25  
|||||  
Db 22060 AAATACTGCACAGAGCCCTCGG 22037

Query Match 70.4%; Score 17.6; DB 3; Length 32881;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAPACGACACAGCGAGCCCTCGG 25  
|||||  
Db 22060 AAATACTGCACAGAGCCCTCGG 22037

;  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-6163

Query Match 68.0%; Score 17; DB 3; Length 603;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25  
||||| ||||| ||||| ||||| |||||  
Db 164 AATAGCGAAACACGCGTCCCTCGG 140

RESULT 4  
US-09-949-016-13114  
; Sequence 13114, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13114  
; LENGTH: 19857  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13114

Query Match 68.0%; Score 17; DB 3; Length 19857;  
Best Local Similarity 80.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25  
||||| ||||| ||||| ||||| |||||  
Db 3315 AAAAAACACACGCCGCCCAAGG 3339

RESULT 5  
US-09-949-016-12657  
; Sequence 12657, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12657  
; LENGTH: 19998  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12657

Query Match 68.0%; Score 17; DB 3; Length 19998;

Best Local Similarity 80.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25  
||||| ||||| ||||| ||||| |||||  
Db 3259 AAAAAACACACGCCGCCCAAGG 3283

RESULT 6  
US-09-949-016-13484  
; Sequence 13484, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13484  
; LENGTH: 36938  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13484

Query Match 68.0%; Score 17; DB 3; Length 36938;  
Best Local Similarity 80.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25  
||||| ||||| ||||| ||||| |||||  
Db 21572 AAGATCCACACAGTGAGCCCTCTG 21596

RESULT 7  
US-09-248-796A-11154/c  
; Sequence 11154, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBIC  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 11154  
; LENGTH: 255  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-11154

Query Match 66.4%; Score 16.6; DB 3; Length 255;  
Best Local Similarity 82.6%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTC 23  
||||| ||||| ||||| ||||| |||||  
Db 250 AAAAAAGACATCGAGCCCCC 228



```
RESULT 8
US-09-949-002-1163/c
; Sequence 1163, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1163
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-1163

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTC 23
Db 56 AAAAAAAGAGCGAGCCCTC 34

RESULT 9
US-09-949-002-7259/c
; Sequence 7259, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7259
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-7259

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTC 23
Db 56 AAAAAAAGAGCGAGCCCTC 34

RESULT 10
US-08-991-677-7/c
; Sequence 7, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Carraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
```

```
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2025
; TYPE: DNA
; ORGANISM: Liquidambar styraciflua
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (60)..(1679)
US-08-991-677-7

Query Match          66.4%; Score 16.6; DB 3; Length 2025;
Best Local Similarity 82.6%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTC 23
Db 1502 AAAAAACGACAGGAGCTCTC 1480

RESULT 11
US-09-949-016-14034/c
; Sequence 14034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14034
; LENGTH: 4493
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14034

Query Match          66.4%; Score 16.6; DB 3; Length 4493;
Best Local Similarity 82.6%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAACGACACAGCGAGCCCTCGG 25
Db 702 AAAAAACACTGCAAGCCCTGGG 680

RESULT 12
US-09-949-002-580/c
; Sequence 580, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

US-09-949-016-16821
Query Match 66.4%; Score 16.6; DB 3; Length 43717;
Best Local Similarity 82.6%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAACGACACAGCGAGCCCTCGG 25
DB 19034 AAAACGACACTGCAAGCCCTCGG 19012

RESULT 15
US-09-300-958A-40
; Sequence 40, Application US/09300958A
; Patent No. 6495319
; GENERAL INFORMATION:
; APPLICANT: McClelland, Michael
; APPLICANT: Welsh, John
; APPLICANT: Trenkle, Thomas
; TITLE OF INVENTION: Reduced Complexity Nucleic Acid Targets and Methods of
; FILE REFERENCE: P-PH 3457
; CURRENT APPLICATION NUMBER: US/09/300,958A
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/083,331
; PRIOR FILING DATE: 1998-04-27
; PRIOR APPLICATION NUMBER: 60/098,070
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: 60/118,624
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 40
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (33)..(34)
; NAME/KEY: unsure
; LOCATION: (59)
; NAME/KEY: unsure
; LOCATION: (82)
; NAME/KEY: unsure
; LOCATION: (112)
; NAME/KEY: unsure
; LOCATION: (184)
; NAME/KEY: unsure
; LOCATION: (126)
; NAME/KEY: unsure
; LOCATION: (164)
; NAME/KEY: unsure
; LOCATION: (184)
; NAME/KEY: unsure
; LOCATION: (225)
; NAME/KEY: unsure
; LOCATION: (244)
; NAME/KEY: unsure
; LOCATION: (253)
; NAME/KEY: unsure
; LOCATION: (272)
; NAME/KEY: unsure
; LOCATION: (307)
; NAME/KEY: unsure
; LOCATION: (316)
; NAME/KEY: unsure
; LOCATION: (329)
; NAME/KEY: unsure
; LOCATION: (335)
; NAME/KEY: unsure
; LOCATION: (381)
; NAME/KEY: unsure
; LOCATION: (396)
; NAME/KEY: unsure
; LOCATION: (417)

US-09-949-016-16821
Query Match 66.4%; Score 16.6; DB 3; Length 40325;
Best Local Similarity 82.6%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGCAGACAGCGAGCCCTC 23
DB 24633 AAAAAGCAGCGAGCCCTC 24611

RESULT 13
US-09-949-002-771/c
; Sequence 771, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 771
; LENGTH: 40325
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(43717)
; OTHER INFORMATION: n = A T C G

US-09-949-016-16821/c
Query Match 66.4%; Score 16.6; DB 3; Length 40325;
Best Local Similarity 82.6%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGCAGACAGCGAGCCCTC 23
DB 24633 AAAAAGCAGCGAGCCCTC 24611

RESULT 14
US-09-949-016-16821/c
; Sequence 16821, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16821
; LENGTH: 43717
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(43717)
; OTHER INFORMATION: n = A T C G

```

```

; NAME/KEY: unsure
; LOCATION: (422)
; NAME/KEY: unsure
; LOCATION: (429)
; NAME/KEY: unsure
; LOCATION: (448)
US-09-300-958A-40

```

Query Match 64.8%; Score 16.2; DB 3; Length 452;  
Best Local Similarity 78.3%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels

QY 1 AAAAAACGACACAGCGAGCCCTC 23  
||||| | ||| |||||  
Db 375 AAAAAAGCCACACTGAGCCNC 397

```

RESULT 16
5242798-6
;Patent No. 5242798
; APPLICANT: SUTCLIFFE, J. GERGOR
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES CORRESPONDING
; TO PORTIONS OF PROTEINOIDS TRANSLATED FROM BRAIN-SPECIFIC MNAS,
; RECEPTORS, METHODS AND DIAGNOSTICS USING THE SAME
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/476,961
; FILING DATE: 07-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 58 620
; FILING DATE: 03-JUN-1987
; APPLICATION NUMBER: 516,136
; FILING DATE: 21-JUL-1983
; SEQ ID NO:6
; LENGTH: 1409
5242798-6

```

Query Match 64.8%; Score 16.2; DB 9; Length 1409;  
Best Local Similarity 85.7%; Pred. No. 5.5e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0

Qy 1 AAAAAACGACACAGCGAGCCC 21  
712 AAAGAACGTACAGAGAGCCC 732

RESULT 17  
 US-09-902-540-3646  
 ; Sequence 3646, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 3646  
 ; LENGTH: 2826  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-3646

Query Match 64.8%; Score 16.2; DB 3; Length 2826;  
Best Local Similarity 85.7%; Pred. No. 5.7e+02;  
Matches 18: Conservative 0; Mismatches 3; Indels 0

QY 5 AACGACACAGCGAGCCCTCGG 25

Db  
1348 AACGACAGGCGGCCCGCG 1368

```

RESULT 18
US-09-902-540-2000
; Sequence 2000, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2000
; LENGTH: 16563
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2000

```

Query Match	64.8%	Score	16.2	DB	3	Length	16563
Best Local Similarity	85.7%	Pred. No.	6.4e+02				
Matches	18	Conservative	0	Mismatches	3	Indels	0
						Gaps	0

Qy 5 AACGACACAGCGAGCCCTCGG 25  
15196 AACGACACAGCGCGCCCGGG 15216

```

RESULT 19
US-09-902-540-1176/c
; Sequence 1176, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1176
; LENGTH: 16844
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1176

```

Query Match	64.8%	Score 16.2;	DB 3;	Length 16844;
Best Local Similarity	85.7%	Pred. No. 6.4e+02;		
Matches 18;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

QY 5 AACGACACAGCGAGCCCTCGG 25  
1481 AACGACACGGCGGCCCGCGG 1461  
Db

RESULT 20  
US-09-902-540-1241/c  
; Sequence 1241, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.

```
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1241
; LENGTH: 28493
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(28493)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1241

Query Match          64.8%; Score 16.2; DB 3; Length 28493;
Best Local Similarity 85.7%; Pred. No. 6.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 AACGACACGCGAGCCCTCGG 25
Db      17737 AACGATACAGCGAGCCACCG 17717

RESULT 21
US-10-024-396-13/c
; Sequence 13, Application US/10024396
; Patent No. 6852536
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF CD36L1 EXPRESSION
; FILE REFERENCE: RTS-0339
; CURRENT APPLICATION NUMBER: US/10/024,396
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 13
; LENGTH: 41322
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon:intron junction
; LOCATION: (9980)...(9981)
; OTHER INFORMATION: exon 8:intron 8
; NAME/KEY: intron
; LOCATION: (17618)...(22472)
; OTHER INFORMATION: intron 9
; NAME/KEY: intron
; LOCATION: (22547)...(30283)
; OTHER INFORMATION: intron 10
; NAME/KEY: intron:exon junction
; LOCATION: (30283)...(30284)
; OTHER INFORMATION: intron 10:exon 11
; NAME/KEY: intron:exon junction
; LOCATION: (31237)...(31238)
; OTHER INFORMATION: intron 11:exon 12
; NAME/KEY: intron
; LOCATION: (31385)...(34929)
; OTHER INFORMATION: intron 12
; NAME/KEY: intron
; LOCATION: (35059)...(39154)
; OTHER INFORMATION: intron 13
; NAME/KEY: intron:exon junction
; LOCATION: (39154)...(39155)
; OTHER INFORMATION: intron 13:exon 14
US-10-024-396-13

Query Match          64.8%; Score 16.2; DB 3; Length 41322;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAACGACACGAGCGCCCTC 23
Db      7922 AACACGACACGAGGCGCTC 7902

RESULT 22
US-09-949-016-14038/c
; Sequence 14038, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14038
; LENGTH: 90472
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(90472)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14038

Query Match          64.8%; Score 16.2; DB 3; Length 90472;
Best Local Similarity 85.7%; Pred. No. 6.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAACGACACGAGCGCCCTC 23
Db      56403 AACACGACACGAGGCGCTC 56383

RESULT 23
US-09-949-016-11989/c
; Sequence 11989, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11989
; LENGTH: 158735
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION:
US-09-949-016-11989

Query Match          64.8%; Score 16.2; DB 3; Length 158735;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAACGACACGAGCGCC 21
Db 55134 ACAAAGACACGACGAGCC 55114

RESULT 24
US-09-949-016-17130/c
; Sequence 17130, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17130
; LENGTH: 158735
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17130

Query Match 64.8%; Score 16.2; DB 3; Length 158735;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCGCC 21
Db 55134 ACAAAGACACGACGAGCC 55114

RESULT 25
US-09-513-999C-16876/c
; Sequence 16876, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 16876
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-16876

Query Match 64.0%; Score 16; DB 3; Length 184;
Best Local Similarity 79.2%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAACGACACGAGCGCCCTCGG 25
Db 102 AAATCGTCACAGACCTCTCGG 79

RESULT 26
US-09-533-559-2241

; Sequence 2241, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Key
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2241
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Fusarium venenatum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(531)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-2241

Query Match 64.0%; Score 16; DB 3; Length 531;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CGACACGAGCGCCCT 22
Db 217 CGACACGAGCGCCCT 232

RESULT 27
US-09-949-016-27169/c
; Sequence 27169, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27169
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27169

Query Match 64.0%; Score 16; DB 3; Length 601;
Best Local Similarity 79.2%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAACGACACGAGCGCCCTCGG 25
Db 144 AAAATGACACAGGCTGCACTCGG 121

RESULT 28
US-09-949-016-178692
```

```
; Sequence 178692, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 178692
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-178692

Query Match          64.0%; Score 16; DB 3; Length 601;
Best Local Similarity 79.2%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
DB 36 ACACAGCAGACAGAGAGCCCGGG 59

RESULT 29
US-09-949-016-178693
; Sequence 178693, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 178693
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-178693

Query Match          64.0%; Score 16; DB 3; Length 601;
Best Local Similarity 79.2%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
DB 536 ACACAGCAGACAGAGAGCCCGGG 559

RESULT 30
US-09-949-016-182367/c
; Sequence 182367, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182367
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-182367

Query Match          64.0%; Score 16; DB 3; Length 601;
Best Local Similarity 79.2%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
DB 144 AAAAATGACACAGCGTGCACCTCGG 121

RESULT 31
US-09-543-681A-2689/c
; Sequence 2689, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2689
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Proteus mirabilis
; US-09-543-681A-2689

Query Match          64.0%; Score 16; DB 3; Length 1347;
Best Local Similarity 79.2%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
DB 876 AATAATGCCACACAGAGCCCGGG 853

RESULT 32
US-09-949-016-651
; Sequence 651, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```

; ORGANISM: Human
; US-09-949-016-16690

Query Match          64.0%; Score 16; DB 3; Length 13970;
Best Local Similarity 79.2%; Pred. No. 7.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCG 24
    ||| ||| ||| ||| ||| ||| ||| |||
Db 3018 AACTACACACAGCAAGCCTTCG 2995

RESULT 35
US-09-949-016-15191/c
; Sequence 15191, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15191
; LENGTH: 27968
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(27968)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15191

Query Match          64.0%; Score 16; DB 3; Length 27968;
Best Local Similarity 79.2%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCG 24
    ||| ||| ||| ||| ||| ||| ||| |||
Db 20778 AAAAAAGCGACAGAGCCCTCG 20755

RESULT 36
US-09-949-016-15192/c
; Sequence 15192, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15192
; LENGTH: 27968
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(27968)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15192

```



```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(27968)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-15192

Query Match      64.0%; Score 16; DB 3; Length 27968;
Best Local Similarity 79.2%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCG 24
   ||||| ||| ||||| ||||| |||||
Db 20778 AAAAAAGGACAGGAGCCCTCG 20755

RESULT 37
US-09-949-016-13098/c
; Sequence 13098, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13098
; LENGTH: 34372
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13098

Query Match      64.0%; Score 16; DB 3; Length 34372;
Best Local Similarity 79.2%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCCTCG 25
   ||||| ||||| ||||| ||||| |||||
Db 3045 AGAAAGGACACAGGAGCCCGG 3022

RESULT 38
US-09-949-016-13099/c
; Sequence 13099, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13099
; LENGTH: 34875
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13099

Query Match      64.0%; Score 16; DB 3; Length 34875;
Best Local Similarity 79.2%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCCTCG 25
   ||||| ||||| ||||| ||||| |||||
Db 3045 AGAAAGGACACAGGAGCCCGG 3022

RESULT 39
US-09-949-016-16872/c
; Sequence 16872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16872
; LENGTH: 37861
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(37861)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16872

Query Match      64.0%; Score 16; DB 3; Length 37861;
Best Local Similarity 79.2%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCCTCG 25
   ||||| ||||| ||||| ||||| |||||
Db 31151 ACACAGACACAGAGCCCGG 31128

RESULT 40
US-09-949-016-12212/c
; Sequence 12212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12212
; LENGTH: 97195
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; US-09-949-016-12212
```

```
; LOCATION: (1)...(97195)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12212

Query Match      64.0%; Score 16; DB 3; Length 97195;
Best Local Similarity 79.2%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCACACGAGCGCCTCGG 25
   ||||| ||||| ||||| ||||| |||||
Db 31799 AAAAATGACACAGGCTGCACTCGG 31776

RESULT 41
US-09-949-016-16971/c
; Sequence 16971, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J, Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16971
; LENGTH: 97196
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(97196)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16971

Query Match      64.0%; Score 16; DB 3; Length 97196;
Best Local Similarity 79.2%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCACACGAGCGCCTCGG 25
   ||||| ||||| ||||| ||||| |||||
Db 31799 AAAAATGACACAGGCTGCACTCGG 31776

RESULT 42
US-08-779-764A-13
; Sequence 13, Application US/08779764A
; Patent No. 6057094
; GENERAL INFORMATION:
; APPLICANT: de la Torre, Juan C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6057094th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: United States
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

Query Match      64.0%; Score 16; DB 3; Length 97196;
Best Local Similarity 79.2%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCACACGAGCGCCTCGG 25
   ||||| ||||| ||||| ||||| |||||
Db 31799 AAAAATGACACAGGCTGCACTCGG 31776

RESULT 43
US-09-563-456-13
; Sequence 13, Application US/09563456
; Patent No. 6653464
; GENERAL INFORMATION:
; APPLICANT: de la Torre, Juan C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6653464th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: United States
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

Query Match      63.2%; Score 15.8; DB 3; Length 571;
Best Local Similarity 73.9%; Pred. No. 7.9e+02;
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGCACACGAGCGCCTC 23
   ||||| ||||| ||||| ||||| |||||
Db 157 AAAAGACACACGCGGTGATCTC 179

US-08-779-764A-13
; APPLICATION NUMBER: US/08/779,764A
; FILING DATE: 16-DEC-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 465.0
; TELEPHONE: (619) 784-2937
; TELEFAX: (619) 784-9399
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-779-764A-13
```

```
; APPLICATION NUMBER: US/08/779,764A
; FILING DATE: 16-DEC-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 465.0
; TELEPHONE: (619) 784-2937
; TELEFAX: (619) 784-9399
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-779-764A-13

Query Match      63.2%; Score 15.8; DB 3; Length 571;
Best Local Similarity 73.9%; Pred. No. 7.9e+02;
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGCACACGAGCGCCTC 23
   ||||| ||||| ||||| ||||| |||||
Db 157 AAAAGACACACGCGGTGATCTC 179

US-09-563-456-13
; Sequence 13, Application US/09563456
; Patent No. 6653464
; GENERAL INFORMATION:
; APPLICANT: de la Torre, Juan C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6653464th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: United States
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/779,764
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 465.0
; TELEPHONE: (619) 784-2937
; TELEFAX: (619) 784-9399
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
```

; ANTI-SENSE: NO  
US-09-563-456-13

Query Match 63.2%; Score 15.8; DB 3; Length 571;  
Best Local Similarity 73.9%; Pred. No. 7.9e+02;  
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACAGCGCCCTC 23  
|||||:|||||:|||||  
Db 157 AAAGACRACRACGGTGATCTC 179

RESULT 44  
US-09-252-991A-7872  
; Sequence 7872, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7872  
; LENGTH: 1692  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7872

Query Match 63.2%; Score 15.8; DB 3; Length 1692;  
Best Local Similarity 89.5%; Pred. No. 8.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 CGACACGCGAGCCCTCGG 25  
|||||:|||||:|||||  
Db 494 CAACCGCGAGCCCTCGG 512

RESULT 45  
US-08-931-999-4  
; Sequence 4, Application US/08931999  
; Patent No. 6043219  
; GENERAL INFORMATION:  
; APPLICANT: Iandolo, John J.  
; APPLICANT: Crupper, Scott S.  
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hovey, Williams, Timmons & Collins  
; STREET: 2405 Grand Boulevard, Suite 400  
; CITY: Kansas City  
; STATE: Missouri  
; COUNTRY: U.S.A.  
; ZIP: 64108  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/931,999  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/710,561  
; FILING DATE: 19-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Collins, John M.

; REGISTRATION NUMBER: 26,262  
; REFERENCE/DOCKET NUMBER: 25043-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 816/474-9050  
; TELEFAX: 816/474-9057  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6755 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus aureus  
; STRAIN: UT0007  
US-08-931-999-4

Query Match 63.2%; Score 15.8; DB 3; Length 6755;  
Best Local Similarity 89.5%; Pred. No. 9.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACAGCGAGC 19  
|||||:|||||:|||||  
Db 2630 AAAAAACGAAACAGCAAGC 2648

RESULT 46  
US-08-779-764A-1  
; Sequence 1, Application US/08779764A  
; Patent No. 6057094  
; GENERAL INFORMATION:  
; APPLICANT: de la Torre, Juan C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING  
; TITLE OF INVENTION: OF HUMAN BORNA DISEASE VIRUS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
; STREET: 10550 No. 6057094th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/779,764A  
; FILING DATE: 16-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSRI 465.0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 784-2937  
; TELEFAX: (619) 784-9399  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-779-764A-1

Query Match 63.2%; Score 15.8; DB 3; Length 8910;  
Best Local Similarity 73.9%; Pred. No. 9.3e+02;

```
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACAGCGAGCCCTC 23
    ||||| :||:||||| |||
Db 415 AAAAGACRACRACGCGTGATCTC 437

RESULT 47
US-09-563-456-1
; Sequence 1, Application US/09563456
; Patent No. 6653464
; GENERAL INFORMATION:
; APPLICANT: de la Torre, Juan C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; TITLE OF INVENTION: OF HUMAN BORNA DISEASE VIRUS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10550 No. 6653464th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: United States
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/563,456
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/779,764
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 465.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 784-2937
; TELEFAX: (619) 784-9399
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-563-456-1

Query Match 63.2%; Score 15.8; DB 3; Length 8910;
Best Local Similarity 73.9%; Pred. No. 9.3e+02;
Matches 17; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACAGCGAGCCCTC 23
    ||||| :||:||||| |||
Db 415 AAAAGACRACRACGCGTGATCTC 437

RESULT 48
US-09-949-016-17073/c
; Sequence 17073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17073
; LENGTH: 72662
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(72662)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17073

Query Match 63.2%; Score 15.8; DB 3; Length 72662;
Best Local Similarity 89.5%; Pred. No. 1.e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 ACGACACAGCGAGCCCTCG 24
    ||||| :||:||||| |||
Db 16179 ACAACACAGCGAGACCTCG 16161

RESULT 49
US-09-949-016-13734
; Sequence 13734, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13734
; LENGTH: 131254
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(131254)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13734

Query Match 63.2%; Score 15.8; DB 3; Length 131254;
Best Local Similarity 89.5%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 ACGACACAGCGAGCCCTCG 24
    ||||| :||:||||| |||
Db 67235 ACGACACAGCGAGCCTTCG 67253

RESULT 50
US-09-949-016-14896
; Sequence 14896, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14896
; LENGTH: 135030
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(135030)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14896

Query Match      63.2%; Score 15.8; DB 3; Length 135030;
Best Local Similarity 89.5%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      5 AACGACACGCGGCGCCTC 23
Db      36869 AACAAACACGCGGACCTC 36887

Search completed: February 3, 2006, 16:32:18
Job time : 86.6667 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-15

Perfect score: 25

Sequence: 1 aaaaacagacacagcgcctcgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	US-10-719-900-15
2	23.4	93.6	25	8	US-10-719-900-16
3	19.8	79.2	770	8	US-10-363-345A-13473
4	19.8	79.2	770	8	US-10-363-345A-13474
5	19.8	79.2	770	9	US-10-363-483A-13473
6	19.8	79.2	770	9	US-10-363-483A-13474
7	18.8	75.2	763	3	US-09-910-943-24
8	18.6	74.4	352	8	US-10-425-115-138645
9	18.6	74.4	378	7	US-10-437-963-28486
10	18.2	72.8	519	5	US-10-027-632-46652
11	18.2	72.8	519	6	US-10-027-632-46652
12	17.8	71.2	1368	3	US-09-974-300-893
13	17.8	71.2	115780	7	US-10-367-094-96
14	17.6	70.4	215	7	US-10-424-599-95630
15	17.6	70.4	559	3	US-09-974-300-6825
16	17.6	70.4	597	3	US-09-974-300-6762
17	17.6	70.4	753	8	US-10-363-345A-26299
18	17.6	70.4	753	8	US-10-363-345A-26300
19	17.6	70.4	753	8	US-10-363-483A-26299
20	17.6	70.4	753	9	US-10-363-483A-26300
21	17.6	70.4	1017	8	US-10-363-345A-28097
22	17.6	70.4	1017	8	US-10-363-345A-28098
23	17.6	70.4	1017	9	US-10-363-483A-28097



C 97	16.6	66.4	936	8	US-10-363-345A-35197	Sequence 35197, A
C 98	16.6	66.4	936	8	US-10-363-345A-35198	Sequence 35198, A
C 99	16.6	66.4	936	8	US-10-363-483A-35197	Sequence 35197, A
C 100	16.6	66.4	936	9	US-10-363-483A-35198	Sequence 35198, A
C 101	16.6	66.4	1114	7	US-10-437-963-93727	Sequence 93727, A
C 102	16.6	66.4	1243	7	US-10-425-114-6511	Sequence 6511, Ap
C 103	16.6	66.4	1252	7	US-10-425-114-35684	Sequence 35684, A
C 104	16.6	66.4	1270	8	US-10-425-115-147902	Sequence 147902, A
C 105	16.6	66.4	1488	10	US-11-097-143-22396	Sequence 22396, A
C 106	16.6	66.4	1645	8	US-10-425-115-12812	Sequence 12812, A
C 107	16.6	66.4	1705	7	US-10-425-114-31800	Sequence 31800, A
C 108	16.6	66.4	1871	8	US-10-425-115-85422	Sequence 85422, A
C 109	16.6	66.4	1896	8	US-10-425-115-147904	Sequence 147904, A
C 110	16.6	66.4	2025	3	US-09-796-256A-7	Sequence 7, Appli
C 111	16.6	66.4	2025	9	US-10-681-878A-7	Sequence 7, Appli
C 112	16.6	66.4	2198	7	US-10-424-599-94282	Sequence 94282, A
C 113	16.6	66.4	31312	7	US-10-322-281-709	Sequence 709, App
C 114	16.6	66.4	81684	7	US-10-322-281-673	Sequence 673, App
C 115	16.6	66.4	106315	5	US-10-087-192-178	Sequence 178, App
C 116	16.6	66.4	439892	5	US-10-087-192-454	Sequence 454, App
C 117	16.6	66.4	1980090	8	US-10-719-993-6815	Sequence 6815, Ap
C 118	16.6	66.4	1980090	8	US-10-741-600-17676	Sequence 17676, A
C 119	16.4	65.6	640	4	US-09-925-065A-668407	Sequence 668407, A
C 120	16.4	65.6	640	4	US-09-925-065A-668408	Sequence 668408, A
C 121	16.4	65.6	723	5	US-10-027-632-17573	Sequence 17573, A
C 122	16.4	65.6	723	5	US-10-027-632-17574	Sequence 17574, A
C 123	16.4	65.6	723	6	US-10-027-632-17573	Sequence 17573, A
C 124	16.4	65.6	723	6	US-10-027-632-17574	Sequence 17574, A
C 125	16.4	65.6	1343	5	US-10-001-843-70	Sequence 70, Appl
C 126	16.4	65.6	1343	10	US-11-005-609-70	Sequence 70, Appl
C 127	16.4	65.6	3259	5	US-10-001-843-71	Sequence 71, Appl
C 128	16.4	65.6	3259	10	US-11-005-609-71	Sequence 71, Appl
C 129	16.2	64.8	25	7	US-10-719-956-1819	Sequence 1819, Ap
C 130	16.2	64.8	259	8	US-10-425-115-32663	Sequence 32663, A
C 131	16.2	64.8	327	8	US-10-425-115-177018	Sequence 177018, A
C 132	16.2	64.8	377	8	US-10-425-115-89755	Sequence 89755, A
C 133	16.2	64.8	452	6	US-10-291-808-40	Sequence 40, Appl
C 134	16.2	64.8	454	4	US-09-925-065A-122779	Sequence 122779, A
C 135	16.2	64.8	454	4	US-09-925-065A-122780	Sequence 122780, A
C 136	16.2	64.8	454	4	US-09-925-065A-840909	Sequence 840909, A
C 137	16.2	64.8	519	7	US-10-437-963-9015	Sequence 9015, Ap
C 138	16.2	64.8	527	4	US-09-925-065A-42101	Sequence 42101, A
C 139	16.2	64.8	529	7	US-10-424-599-12785	Sequence 12785, A
C 140	16.2	64.8	542	4	US-09-925-065A-428599	Sequence 428599, A
C 141	16.2	64.8	571	8	US-10-363-345A-34329	Sequence 34329, A
C 142	16.2	64.8	571	8	US-10-363-345A-34330	Sequence 34330, A
C 143	16.2	64.8	571	9	US-10-363-483A-34329	Sequence 34329, A
C 144	16.2	64.8	571	9	US-10-363-483A-34330	Sequence 34330, A
C 145	16.2	64.8	599	9	US-10-972-079-66015	Sequence 66015, A
C 146	16.2	64.8	600	9	US-10-972-079-66016	Sequence 66016, A
C 147	16.2	64.8	600	9	US-10-972-079-66017	Sequence 66017, A
C 148	16.2	64.8	600	9	US-10-972-079-66018	Sequence 66018, A
C 149	16.2	64.8	600	9	US-10-972-079-66019	Sequence 66019, A
C 150	16.2	64.8	607	8	US-10-363-345A-34775	Sequence 34775, A

RESULT 1

US-10-719-900-15

;

SEQ ID NO 15

;

;

LENGTH: 25

;

;

TYPE: DNA

;

;

ORGANISM: Mus musculus

;

US-10-719-900-15

;

Query Match

100.0%; Score 25; DB 8; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.099;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1

AAAAAACGACACAGCGAGCCCTCGG 25

|||||

Db

1

AAAAAACGACACAGCGAGCCCTCGG 25

|||||

RESULT 2

US-10-719-900-16

;

Sequence 16, Application US/10719900

;

;

Publication No. US20050026164A1

;

;

GENERAL INFORMATION:

;

;

APPLICANT: Xue Mei Zhou

;

;

TITLE OF INVENTION: Methods of Genetic Analysis of Mouse

;

;

FILE REFERENCE: 3528.1

;

;

CURRENT APPLICATION NUMBER: US/10719,900

;

;

CURRENT FILING DATE: 2003-11-20

;

;

PRIOR APPLICATION NUMBER: 60/427,808

;

;

PRIOR FILING DATE: 2002 11 20

;

;

NUMBER OF SEQ ID NOS: 982914

;

;

SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

;

;

SEQ ID NO 16

;

;

LENGTH: 25

;

;

TYPE: DNA

;

;

ORGANISM: Mus musculus

;

US-10-719-900-16

;

Query Match

93.6%; Score 23.4; DB 8; Length 25;

Best Local Similarity 96.0%; Pred. No. 0.55;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1

AAAAAACGACACAGCGAGCCCTCGG 25

|||||

Db

1

AAAAAACGACACTGGCGAGCCCTCGG 25

|||||

RESULT 3

US-10-363-345A-13473/c

;

Sequence 13473, Application US/10363345A

;

;

Publication No. US20040234960A1

;

;

GENERAL INFORMATION:

;

;

APPLICANT: Alexander Olek

;

;

APPLICANT: Christian Piepenbrock

;

;

APPLICANT: Kurt Berlin

;

;

TITLE OF INVENTION: Method for determining the degree of methylation of defined

;

;

TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3

;

;

FILE REFERENCE: E01/1237

;

;

CURRENT APPLICATION NUMBER: US/10/363,345A

;

;

CURRENT FILING DATE: 2003-03-03

;

;

NUMBER OF SEQ ID NOS: 40712

;

;

SEQ ID NO 13473

;

;

LENGTH: 770

;

;

TYPE: DNA

;

;

ORGANISM: Artificial Sequence

;

;

FEATURE:

;

;

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

;

;

OTHER INFORMATION: CpG-island NO: 13473

;

US-10-363-345A-13473

;

Query Match

79.2%; Score 19.8; DB 8; Length 770;

Best Local Similarity 91.3%; Pred. No. 29;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1

AAAAAACGACACAGCGAGCCCTC 23

|||||

Db 409 AAAAAACGACACCGCGACCCCTC 387

## RESULT 4

US-10-363-345A-13474  
; Sequence 13474, Application US/10363345A  
; Publication No. US20040234960A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Method for determining the degree of methylation of defined  
; TITLE OF INVENTION: Cytosines in genomic DNA in the sequence context of 5'-CpG-3  
; FILE REFERENCE: E01/1227  
; CURRENT APPLICATION NUMBER: US/10/363.345A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 13474  
; LENGTH: 770  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 13474  
US-10-363-345A-13474

Query Match 79.2%; Score 19.8; DB 8; Length 770;  
Best Local Similarity 91.3%; Pred. No. 29;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACCGCGACCCCTC 23

Db 362 AAAAAACGACACCGCGACCCCTC 384

## RESULT 5

US-10-363-483A-13473/c  
; Sequence 13473, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; TITLE OF INVENTION: illnesses  
; FILE REFERENCE: 82011  
; CURRENT APPLICATION NUMBER: US/10/363.483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 13473  
; LENGTH: 770  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 13473  
US-10-363-483A-13473

Query Match 79.2%; Score 19.8; DB 9; Length 770;  
Best Local Similarity 91.3%; Pred. No. 29;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACCGCGACCCCTC 23

Db 409 AAAAAACGACACCGCGACCCCTC 387

## RESULT 6

US-10-363-483A-13474  
; Sequence 13474, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek

; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; TITLE OF INVENTION: illnesses  
; FILE REFERENCE: 82011  
; CURRENT APPLICATION NUMBER: US/10/363.483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 13474  
; LENGTH: 770  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 13474  
US-10-363-483A-13474

Query Match 79.2%; Score 19.8; DB 9; Length 770;  
Best Local Similarity 91.3%; Pred. No. 29;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACCGCGACCCCTC 23

Db 362 AAAAAACGACACCGCGACCCCTC 384

## RESULT 7

US-09-910-943-24  
; Sequence 24, Application US/09910943  
; Patent No. US20020081610A1  
; GENERAL INFORMATION:  
; APPLICANT: Hemmati-Brivanlou, Ali  
; APPLICANT: Altman, Curtis  
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression  
; FILE REFERENCE: 7529/1G148US1  
; CURRENT APPLICATION NUMBER: US/09/910.943  
; CURRENT FILING DATE: 2001-07-23  
; NUMBER OF SEQ ID NOS: 742  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Xenopus laevis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(763)  
; OTHER INFORMATION: n may be a or g or c or t/u  
US-09-910-943-24

Query Match 75.2%; Score 18.8; DB 3; Length 763;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AAACGACACGACCGACCCCTCGG 25

Db 549 AAACGACACGACGAGAACCCCTCGG 570

## RESULT 8

US-10-425-115-138645  
; Sequence 138645, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425.115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326



US-09-974-300-893/c  
; Sequence 893, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085, 500-US  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 893  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-893

Query Match 71.2%; Score 17.8; DB 3; Length 1368;  
Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAACGACACGAGCCCT 22  
|||||  
DB 1250 AAAAAACGACAAAGGAGCCCT 1230

RESULT 13  
US-10-367-094-96/c  
; Sequence 96, Application US/10367094  
; Publication No. US20040170982A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc Malandro  
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
; FILE REFERENCE: 529452001500  
; CURRENT APPLICATION NUMBER: US/10/367,094  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 203  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 96  
; LENGTH: 115780  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(115780)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-367-094-96

Query Match 71.2%; Score 17.8; DB 7; Length 115780;  
Best Local Similarity 90.5%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACGACACGAGCCCTCGG 25  
|||||  
DB 111059 AAGGACAAAGGAGCCCTCGG 111039

RESULT 14  
US-10-424-599-95630/c  
; Sequence 95630, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 95630  
; LENGTH: 215  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_57365C.1  
US-10-424-599-95630

Query Match 70.4%; Score 17.6; DB 7; Length 215;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
DB 32 AAAAAACGACCAAGTGCCTCG 9

RESULT 15  
US-09-974-300-6825  
; Sequence 6825, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085, 500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6825  
; LENGTH: 559  
; TYPE: DNA  
; ORGANISM: Bacillus clausii  
US-09-974-300-6825

Query Match 70.4%; Score 17.6; DB 3; Length 559;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
DB 203 AAAAAACGACATCGAAGCTTCG 226

RESULT 16  
US-09-974-300-6762  
; Sequence 6762, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085, 500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6762  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Bacillus clausii  
US-09-974-300-6762

Query Match 70.4%; Score 17.6; DB 3; Length 597;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 243 AAAAAACGACATCGAAGCTTCG 266

RESULT 17  
US-10-363-345A-26299/c  
; Sequence 26299, Application US/10363345A  
; Publication No. US20040234960A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Method for determining the degree of methylation of defined  
; FILE REFERENCE: E01/1227  
; CURRENT APPLICATION NUMBER: US/10/363,345A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 26299  
; LENGTH: 753  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 26299  
US-10-363-345A-26299

Query Match 70.4%; Score 17.6; DB 8; Length 753;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 468 AAAAAACGACCCGCGACCCCGCG 445

RESULT 18  
US-10-363-345A-26300  
; Sequence 26300, Application US/10363345A  
; Publication No. US20040234960A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Method for determining the degree of methylation of defined  
; FILE REFERENCE: E01/1227  
; CURRENT APPLICATION NUMBER: US/10/363,345A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 26300  
; LENGTH: 753  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 26300  
US-10-363-345A-26300

Query Match 70.4%; Score 17.6; DB 8; Length 753;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 286 AAAAAACGACCCGCGACCCCGCG 309

RESULT 19  
US-10-363-483A-26299/c  
; Sequence 26299, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; FILE REFERENCE: 82011  
; CURRENT APPLICATION NUMBER: US/10/363,483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 26299  
; LENGTH: 753  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 26299  
US-10-363-483A-26299

Query Match 70.4%; Score 17.6; DB 9; Length 753;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 468 AAAAAACGACCCGCGACCCCGCG 445

RESULT 20  
US-10-363-483A-26300  
; Sequence 26300, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; APPLICANT: Kurt Berlin  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; FILE REFERENCE: 82011  
; CURRENT APPLICATION NUMBER: US/10/363,483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 26300  
; LENGTH: 753  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 26300  
US-10-363-483A-26300

Query Match 70.4%; Score 17.6; DB 9; Length 753;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGAGCCCTCG 24  
|||||  
Db 286 AAAAAACGACCCGCGACCCCGCG 309

RESULT 21  
US-10-363-345A-28097/c

; Sequence 28097, Application US/10363345A  
; Publication No. US20040234960A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; TITLE OF INVENTION: Method for determining the degree of methylation of defined  
; FILE REFERENCE: E01/1227  
; CURRENT APPLICATION NUMBER: US/10/363,345A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 28097  
; LENGTH: 1017  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 28097  
US-10-363-345A-28097

Query Match 70.4%; Score 17.6; DB 8; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCG 24  
|||||  
DB 134 AAAAAACGACACGCGTACCCACG 111

## RESULT 22

US-10-363-345A-28098  
; Sequence 28098, Application US/10363345A  
; Publication No. US20040234960A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; TITLE OF INVENTION: Method for determining the degree of methylation of defined  
; FILE REFERENCE: E01/1227  
; CURRENT APPLICATION NUMBER: US/10/363,345A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 28098  
; LENGTH: 1017  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 28098  
US-10-363-345A-28098

Query Match 70.4%; Score 17.6; DB 8; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCG 24  
|||||  
DB 884 AAAAAACGACACGCGTACCCACG 907

## RESULT 23

US-10-363-483A-28097/c  
; Sequence 28097, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; FILE REFERENCE: 82011

; CURRENT APPLICATION NUMBER: US/10/363,483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 28097  
; LENGTH: 1017  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 28097  
US-10-363-483A-28097

Query Match 70.4%; Score 17.6; DB 9; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCG 24  
|||||  
DB 134 AAAAAACGACACGCGTACCCACG 111

## RESULT 24

US-10-363-483A-28098  
; Sequence 28098, Application US/10363483A  
; Publication No. US20050064401A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Olek  
; APPLICANT: Christian Piepenbrock  
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain  
; FILE REFERENCE: 82011  
; CURRENT APPLICATION NUMBER: US/10/363,483A  
; CURRENT FILING DATE: 2003-03-03  
; NUMBER OF SEQ ID NOS: 40712  
; SEQ ID NO 28098  
; LENGTH: 1017  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; OTHER INFORMATION: CpG-island No: 28098  
US-10-363-483A-28098

Query Match 70.4%; Score 17.6; DB 9; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCCTCG 24  
|||||  
DB 884 AAAAAACGACACGCGTACCCACG 907

## RESULT 25

US-10-425-115-101462  
; Sequence 101462, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 101462  
; LENGTH: 1445  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:

```
; OTHER INFORMATION: Clone ID: MRT4577_24045C.1
US-10-425-115-101462

Query Match      70.4%; Score 17.6; DB 8; Length 1445;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACGACACGCGGCCCTCG 25
   ||| ||||| ||||| ||||| |||||
Db 48 AAAGACGACACGCGGCCGCG 71

RESULT 26
US-10-357-930-22662
; Sequence 22662, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22662
; LENGTH: 6073
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6067, 6068, 6069, 6070, 6071, 6072, 6073
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-22662

Query Match      70.4%; Score 17.6; DB 8; Length 6073;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGGCCCTCG 24
   ||||| ||| ||| ||||| |||||
Db 12 AAAAAAGAGACTCGCGGCCCTCG 35

RESULT 27
US-10-357-930-28508
; Sequence 28508, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22662
; LENGTH: 6073
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6067, 6068, 6069, 6070, 6071, 6072, 6073
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-22662

Query Match      70.4%; Score 17.6; DB 8; Length 6073;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGGCCCTCG 24
   ||||| ||| ||| ||||| |||||
Db 12 AAAAAAGAGACTCGCGGCCCTCG 35

RESULT 28
US-09-925-065A-293623/c
; Sequence 293623, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293623
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(597)
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-293623

Query Match      68.8%; Score 17.2; DB 4; Length 597;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```



QY 1 AAAAAACGACACGAGCGCCT 22  
|||||  
Db 580 AAAAAACACAAAGCGAGCACT 559  
|||||  
RESULT 29  
US-09-925-065A-293624/c  
; Sequence 293624, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108927.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 293624  
; LENGTH: 1092  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1092)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-925-065A-293624

Query Match 68.8%; Score 17.2; DB 4; Length 1092;  
Best Local Similarity 86.4%; Pred. No. 4.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCT 22  
|||||  
Db 580 AAAAAACACAAAGCGAGCACT 559  
|||||  
RESULT 30  
US-10-369-493-43475  
; Sequence 43475, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 43475  
; LENGTH: 1797  
; TYPE: DNA  
; ORGANISM: No. US20030233675Altoc punctiforme  
US-10-369-493-43475  
Query Match 68.8%; Score 17.2; DB 6; Length 1797;

Best Local Similarity 86.4%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCT 22  
|||||  
Db 336 AAAAAACGAAACAGCGATCGCT 357  
|||||  
RESULT 31  
US-10-221-596B-22/c  
; Sequence 22, Application US/10221596B  
; Publication No. US20040086855A1  
; GENERAL INFORMATION:  
; APPLICANT: Iwamoto, Masao  
; APPLICANT: Higo, Kenichi  
; TITLE OF INVENTION: Method for Screening Genes Expressing at Desired Sites  
; FILE REFERENCE: 59150-8020  
; CURRENT APPLICATION NUMBER: US/10/221,596B  
; CURRENT FILING DATE: 2003-10-24  
; PRIOR APPLICATION NUMBER: PCT/JP01/10195  
; PRIOR FILING DATE: 2001-11-21  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 6240  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: AP000559  
; DATABASE ENTRY DATE: OCT-1999  
; RELEVANT RESIDUES: 7921 TO 14161  
US-10-221-596B-22

Query Match 68.8%; Score 17.2; DB 7; Length 6240;  
Best Local Similarity 86.4%; Pred. No. 5.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCT 22  
|||||  
Db 1786 AAAAAATGACACAGCAAGCCAT 1765  
|||||  
RESULT 32  
US-10-741-600-17573  
; Sequence 17573, Application US/10741600  
; Publication No. US20050026169A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001499  
; CURRENT APPLICATION NUMBER: US/10/741,600  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 73997  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17573  
; LENGTH: 29686  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(29686)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-741-600-17573  
Query Match 68.8%; Score 17.2; DB 8; Length 29686;  
Best Local Similarity 86.4%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAACGACACGAGCGCCT 22  
|||||  
Db 26855 AAAAAATGACAAAGCTAGCCCT 26876  
|||||

RESULT 33  
US-10-424-599-19441  
; Sequence 19441, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 19441  
; LENGTH: 447  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_117559C.1  
US-10-424-599-19441

Query Match 68.0%; Score 17; DB 7; Length 447;  
Best Local Similarity 80.0%; Pred. No. 5.8e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 45 AAAAAACGACACAGCGAGCCCTCGG 69

RESULT 34  
US-10-425-115-10108  
; Sequence 10108, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 10108  
; LENGTH: 493  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(493)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_10921C.1  
US-10-425-115-10108

Query Match 68.0%; Score 17; DB 8; Length 493;  
Best Local Similarity 80.0%; Pred. No. 5.8e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 261 AAAAAACGACACAGCGAGCCCTCGG 285

RESULT 35  
US-10-663-561-333  
; Sequence 333, Application US/10663561  
; Publication No. US20040166509A1

; GENERAL INFORMATION:  
; APPLICANT: DENSLOW, NANCY D.  
; APPLICANT: LARKIN, PATRICK M.  
; APPLICANT: FOLMAR, LEROY C.  
; APPLICANT: HEMMER, MICHAEL J.  
; TITLE OF INVENTION: DETECTING HORMONALLY ACTIVE COMPOUNDS  
; FILE REFERENCE: 5853-238  
; CURRENT APPLICATION NUMBER: US/10/663,561  
; CURRENT FILING DATE: 2003-09-15  
; NUMBER OF SEQ ID NOS: 566  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 333  
; LENGTH: 602  
; TYPE: DNA  
; ORGANISM: Cyprinodon variegatus  
US-10-663-561-333

Query Match 68.0%; Score 17; DB 7; Length 602;  
Best Local Similarity 80.0%; Pred. No. 5.9e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 547 AAAAAACGACACAGCGAGCCCTCGG 571

RESULT 36  
US-10-027-632-189552/c  
; Sequence 189552, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 189552  
; LENGTH: 615  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-189552

Query Match 68.0%; Score 17; DB 5; Length 615;  
Best Local Similarity 80.0%; Pred. No. 5.9e+02; Mismatches 5; Indels 0; Gaps 0;  
Matches 20; Conservative 0

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25  
|||||  
Db 569 AAAAAAGGACACAGCGAGCCGCGG 545

RESULT 37  
US-10-027-632-189553/c  
; Sequence 189553, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 189553  
; LENGTH: 615  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-189553

Query Match 68.0%; Score 17; DB 5; Length 615;  
Best Local Similarity 80.0%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 569 AGAAAAGGACACAGCGAGCCCGCGG 545

RESULT 38  
US-10-027-632-189552/c  
; Sequence 189552, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 189552  
; LENGTH: 615  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-189552

Query Match 68.0%; Score 17; DB 6; Length 615;  
Best Local Similarity 80.0%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 569 AGAAAAGGACACAGCGAGCCCGCGG 545

RESULT 39  
US-10-027-632-189553/c  
; Sequence 189553, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 189553  
; LENGTH: 615  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-189553

Query Match 68.0%; Score 17; DB 6; Length 615;  
Best Local Similarity 80.0%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAACGACACAGCGAGCCCTCGG 25  
Db 569 AGAAAAGGACACAGCGAGCCCGCGG 545

RESULT 40  
US-09-925-065A-803090/c  
; Sequence 803090, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 803090  
; LENGTH: 619



```
RESULT 44
US-10-027-632-310475
; Sequence 310475, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 310475
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-310475

Query Match          68.0%; Score 17; DB 5; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  AAAAAACGACACAGCGAGCCCTCGG 25
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      355 AAAGATCCACACAGTGGCCCTCTG 379

RESULT 45
US-10-027-632-43803
; Sequence 43803, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43803
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-43803

Query Match          68.0%; Score 17; DB 5; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  AAAAAACGACACAGCGAGCCCTCGG 25
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      355 AAAGATCCACACAGTGGCCCTCTG 379

RESULT 46
US-10-027-632-63884
; Sequence 63884, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63884
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-63884

Query Match          68.0%; Score 17; DB 6; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  AAAAAACGACACAGCGAGCCCTCGG 25
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      355 AAAGATCCACACAGTGGCCCTCTG 379

RESULT 47
US-10-027-632-179577
; Sequence 179577, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43803
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-43803

Query Match          68.0%; Score 17; DB 6; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  AAAAAACGACACAGCGAGCCCTCGG 25
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      355 AAAGATCCACACAGTGGCCCTCTG 379

RESULT 46
US-10-027-632-63884
; Sequence 63884, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63884
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-63884

Query Match          68.0%; Score 17; DB 6; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  AAAAAACGACACAGCGAGCCCTCGG 25
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      355 AAAGATCCACACAGTGGCCCTCTG 379

RESULT 47
US-10-027-632-179577
; Sequence 179577, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

```
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179577
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-179577

Query Match      68.0%; Score 17; DB 6; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
   ||| ||| ||| ||| ||| ||| ||| |||
DB 355 AAAGATCCACACAGTGAGCCCTCTG 379

RESULT 48
US-10-027-632-310475
; Sequence 310475, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 310475
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(646)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-310475

Query Match      68.0%; Score 17; DB 6; Length 646;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
   ||| ||| ||| ||| ||| ||| ||| |||
DB 355 AAAGATCCACACAGTGAGCCCTCTG 379
```

```
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
   ||| ||| ||| ||| ||| ||| ||| |||
DB 355 AAAGATCCACACAGTGAGCCCTCTG 379

RESULT 49
US-10-450-763-825
; Sequence 825, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790C1F3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 825
; LENGTH: 6092
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (6092)..(6069)
; OTHER INFORMATION: 100% homologous to Homo sapiens Polypeptide identified by the
; OTHER INFORMATION: signal sequence trap method, accession number Y02377, Smith-Waterman
; OTHER INFORMATION: Score=44.
US-10-450-763-825

Query Match      68.0%; Score 17; DB 9; Length 6092;
Best Local Similarity 80.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
   ||| ||| ||| ||| ||| ||| ||| |||
DB 2200 AAAAAAGGAGATAGCAACCCCTCAG 2224

RESULT 50
US-09-764-855-247
; Sequence 247, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 17431
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-764-855-247

Query Match      68.0%; Score 17; DB 3; Length 17431;
Best Local Similarity 80.0%; Pred. No. 6.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCTCGG 25
   ||| ||| ||| ||| ||| ||| ||| |||
DB 7370 AAAAAACAAATGCGAGCCCTCTG 7394
```

Search completed: February 3, 2006, 15:43:46  
Job time : 370.556 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-15

Perfect score: 25

Sequence: 1 aaaaacgacacagcgcctcgg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 41903697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New:

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.8	71.2	2347	7	US-10-510-386-23
C 2	17.2	68.8	172649	8	US-11-121-086-36
C 3	17	68.0	860	7	US-10-750-185-51708
C 4	17	68.0	860	7	US-10-750-623-51708
C 5	17	68.0	3555	7	US-10-750-185-31750
C 6	17	68.0	3555	7	US-10-750-623-31750
C 7	17	68.0	198161	7	US-10-775-169-52
C 8	16.6	66.4	1177	7	US-10-750-185-40395
C 9	16.6	66.4	1177	7	US-10-750-623-40395
C 10	16.6	66.4	1524	7	US-10-750-185-44232
C 11	16.6	66.4	1524	7	US-10-750-623-44232
C 12	16.2	64.8	201	8	US-11-124-368A-3989
C 13	16	64.0	201	7	US-10-995-561-72930
C 14	16	64.0	1409	7	US-10-750-185-28365
C 15	16	64.0	1409	7	US-10-750-623-28365
C 16	16	64.0	1432	8	US-11-000-688-716
C 17	16	64.0	1505	7	US-10-750-185-36115
C 18	16	64.0	1505	7	US-10-750-623-36115
C 19	16	64.0	60754	7	US-10-995-561-13440
C 20	16	64.0	96128	7	US-10-995-561-13197
C 21	16	64.0	130733	8	US-11-121-086-19
C 22	16	64.0	162085	8	US-11-121-086-7
C 23	16	64.0	164810	8	US-11-121-086-4
C 24	15.8	63.2	25	8	US-11-121-849-379393
C 25	15.8	63.2	25	8	US-11-136-527-352769
C 26	15.8	63.2	1400	8	US-11-136-527-7464
C 27	15.8	63.2	1713	7	US-10-750-185-54749
C 28	15.8	63.2	1713	7	US-10-750-623-54749
C 29	15.8	63.2	2383	8	US-11-136-527-3368
C 30	15.6	62.4	201	7	US-10-995-561-62101
C 31	15.6	62.4	201	7	US-10-995-561-62151
C 32	15.6	62.4	777	7	US-10-750-185-58178
C 33	15.6	62.4	777	7	US-10-750-623-58178
C 34	15.6	62.4	794	7	US-10-750-185-29381
C 35	15.6	62.4	794	7	US-10-750-623-29381
C 36	15.6	62.4	870	8	US-11-041-776-65
C 37	15.6	62.4	1045	7	US-10-750-185-59482
C 38	15.6	62.4	1045	7	US-10-750-623-59482
C 39	15.6	62.4	1197	8	US-11-136-527-132
C 40	15.6	62.4	1490	7	US-10-750-185-44586
C 41	15.6	62.4	1490	7	US-10-750-623-44586
C 42	15.6	62.4	1700	7	US-10-750-185-52468
C 43	15.6	62.4	1700	7	US-10-750-623-52468
C 44	15.6	62.4	1771	7	US-10-750-185-35903
C 45	15.6	62.4	1771	7	US-10-750-623-35903
C 46	15.6	62.4	54767	7	US-10-995-561-13357
C 47	15.6	62.4	151169	8	US-11-121-086-38
C 48	15.4	61.6	55	6	US-10-517-151-105
C 49	15.4	61.6	121	7	US-10-310-914A-845
C 50	15.4	61.6	201	7	US-10-995-561-41053
C 51	15.4	61.6	201	7	US-10-995-561-41056
C 52	15.4	61.6	711	7	US-10-750-185-43072
C 53	15.4	61.6	711	7	US-10-750-623-43072
C 54	15.4	61.6	1060	7	US-10-750-185-58209
C 55	15.4	61.6	1060	7	US-10-750-623-58209
C 56	15.4	61.6	1083	7	US-10-750-185-27234
C 57	15.4	61.6	1083	7	US-10-750-623-27234
C 58	15.4	61.6	1385	7	US-10-750-185-28881
C 59	15.4	61.6	1385	7	US-10-750-623-28881
C 60	15.4	61.6	1734	7	US-10-750-185-37863
C 61	15.4	61.6	1734	7	US-10-750-623-37863
C 62	15.4	61.6	1930	7	US-10-750-185-47014
C 63	15.4	61.6	1930	7	US-10-750-623-47014
C 64	15.4	61.6	2080	7	US-10-750-185-48927
C 65	15.4	61.6	2080	7	US-10-750-623-48927
C 66	15.4	61.6	2626	7	US-10-750-185-58314
C 67	15.4	61.6	2626	7	US-10-750-623-58314
C 68	15.4	61.6	4589	7	US-10-750-185-62517
C 69	15.4	61.6	4589	7	US-10-750-623-62517
C 70	15.4	61.6	4887	7	US-10-750-185-61589
C 71	15.4	61.6	4887	7	US-10-750-623-61589
C 72	15.4	61.6	15042	7	US-10-995-561-13391
C 73	15.4	61.6	86081	7	US-10-995-561-13246
C 74	15.4	61.6	161994	8	US-11-112-908-57
C 75	15.4	61.6	167891	8	US-11-121-086-14
C 76	15.4	61.6	191091	8	US-11-121-086-60
C 77	15.4	61.6	197781	8	US-11-112-908-34
C 78	15.4	61.6	321019	7	US-10-995-561-13204
C 79	15.2	60.8	25	8	US-11-121-849-199027
C 80	15.2	60.8	455	8	US-11-000-688-1532
C 81	15.2	60.8	3119	8	US-11-000-688-398
C 82	15.2	60.8	5798	8	US-11-136-527-3895
C 83	15.2	60.8	20492	8	US-11-095-668-2
C 84	15.2	60.8	127917	7	US-10-775-169-82
C 85	15.2	60.8	1080000	7	US-10-928-446A-1
C 86	15.2	60.8	1080000	7	US-10-928-446A-181
C 87	15.2	60.8	1080000	7	US-10-928-446A-183
C 88	15.2	60.8	1080000	7	US-10-928-446A-185
C 89	15.2	60.8	1080000	7	US-10-928-446A-187
C 90	15.2	60.8	1080000	7	US-10-928-446A-189
C 91	15.2	60.8	1080000	7	US-10-928-446A-191
C 92	15.2	60.8	1080000	7	US-10-928-446A-193
C 93	15.2	60.8	1080000	7	US-10-928-446A-195
C 94	15.2	60.8	1080000	7	US-10-928-446A-197
C 95	15.2	60.8	1080000	7	US-10-928-446A-199

Sequence 4, Appli  
Sequence 379393,  
Sequence 352769,  
Sequence 7464, Ap  
Sequence 54749, A  
Sequence 54749, A  
Sequence 3368, Ap  
Sequence 62101, A  
Sequence 62151, A  
Sequence 58178, A  
Sequence 58178, A  
Sequence 29381, A  
Sequence 29381, A  
Sequence 59482, A  
Sequence 59482, A  
Sequence 132, App  
Sequence 44586, A  
Sequence 44586, A  
Sequence 52468, A  
Sequence 52468, A  
Sequence 35903, A  
Sequence 35903, A  
Sequence 13357, A  
Sequence 38, Appl  
Sequence 105, App  
Sequence 845, App  
Sequence 41053, A  
Sequence 41056, A  
Sequence 43072, A  
Sequence 43072, A  
Sequence 58209, A  
Sequence 58209, A  
Sequence 27234, A  
Sequence 27234, A  
Sequence 27234, A  
Sequence 28881, A  
Sequence 28881, A  
Sequence 37863, A  
Sequence 37863, A  
Sequence 47014, A  
Sequence 47014, A  
Sequence 48927, A  
Sequence 48927, A  
Sequence 58314, A  
Sequence 58314, A  
Sequence 62517, A  
Sequence 62517, A  
Sequence 62517, A  
Sequence 61589, A  
Sequence 61589, A  
Sequence 13391, A  
Sequence 13246, A  
Sequence 57, Appl  
Sequence 14, Appl  
Sequence 60, Appl  
Sequence 34, Appl  
Sequence 13204, A  
Sequence 13204, A  
Sequence 1532, Ap  
Sequence 388, App  
Sequence 3895, Ap  
Sequence 2, Appl  
Sequence 82, Appl  
Sequence 1, Appl  
Sequence 181, App  
Sequence 183, App  
Sequence 185, App  
Sequence 187, App  
Sequence 189, App  
Sequence 191, App  
Sequence 193, App  
Sequence 195, App  
Sequence 197, App  
Sequence 199, App



```
; TYPE: DNA
; ORGANISM: Bovine 19866881570598
US-10-750-185-51708

Query Match      68.0%; Score 17; DB 7; Length 860;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25
Db 728 AAAAAAGTCACAGCAATCCCTCTG 752

RESULT 4
US-10-750-623-51708
; Sequence 51708, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51708
; LENGTH: 860
; TYPE: DNA
; ORGANISM: Bovine 19866881570598
US-10-750-623-51708

Query Match      68.0%; Score 17; DB 7; Length 860;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25
Db 728 AAAAAAGTCACAGCAATCCCTCTG 752

RESULT 5
US-10-750-185-31750/c
; Sequence 31750, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31750
; LENGTH: 3555
; TYPE: DNA
; ORGANISM: Bovine 19866881076987
US-10-750-185-31750

Query Match      68.0%; Score 17; DB 7; Length 860;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25
Db 728 AAAAAAGTCACAGCAATCCCTCTG 752

RESULT 6
US-10-750-623-31750/c
; Sequence 31750, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31750
; LENGTH: 3555
; TYPE: DNA
; ORGANISM: Bovine 19866881076987
US-10-750-623-31750

Query Match      68.0%; Score 17; DB 7; Length 3555;
Best Local Similarity 80.0%; Pred. No. 81;
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25
Db 3525 AAGAAACGCGAGAGCAAGCCCTCGG 3501

RESULT 7
US-10-775-169-52/c
; Sequence 52, Application US/10775169
; Publication No. US20050287532A9
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dörner, Andrew
; APPLICANT: Trepicchio, William
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
; FILE REFERENCE: AM101080 (031896-013000)
; CURRENT APPLICATION NUMBER: US/10775,169
; CURRENT FILING DATE: 2004-02-11
; NUMBER OF SEQ ID NOS: 5278
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 52
; LENGTH: 198161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-169-52

Query Match      68.0%; Score 17; DB 7; Length 198161;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Qy 1 AAAAAACGACACGAGCGCCCTCGG 25
Db 198161 AAAAAAGTCACAGCAATCCCTCTG 198161
```



```
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3989
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-3989

Query Match 64.8%; Score 16.2; DB 8; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAGCAGACGAGCCCT 22
Db 67 AAAAAGCAGACGAGCCCT 87

RESULT 13
US-10-995-561-72930/c
; Sequence 72930, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72930
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-72930

Query Match 64.0%; Score 16; DB 7; Length 201;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAAAAGCAGACGAGCCCTCGG 25
Db 85 AAAAAGCAGACGAGCGGCACGCTG 62

RESULT 14
US-10-750-185-28365/c
; Sequence 28365, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

```
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28365
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Bovine 19866881382047
US-10-750-185-28365

Query Match 64.0%; Score 16; DB 7; Length 1409;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAAAAGCAGACGAGCCCTCGG 25
Db 64 AGAAAGAAACAGCAAGCCCTCGG 41

RESULT 15
US-10-750-623-28365/c
; Sequence 28365, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28365
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Bovine 19866881382047
US-10-750-623-28365

Query Match 64.0%; Score 16; DB 7; Length 1409;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAAAAGCAGACGAGCCCTCGG 25
Db 64 AGAAAGAAACAGCAAGCCCTCGG 41

RESULT 16
US-11-000-688-716
; Sequence 716, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 716
; LENGTH: 1432
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; NAME/KEY: misc feature
; LOCATION: (1)..(1432)
; OTHER INFORMATION: phosphorylase kinase, gamma 1
; OTHER INFORMATION: (muscle)(PHKG1) gene.
US-11-000-688-716

Query Match          64.0%; Score 16; DB 8; Length 1432;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
Db 154 AGAATGACAGCGGAGCCATGGG 177

RESULT 17
US-10-750-185-36115
; Sequence 36115, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36115
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Bovine 19866880565425
US-10-750-185-36115

Query Match          64.0%; Score 16; DB 7; Length 1505;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
Db 898 AAACAGACACATGAGGCTAGG 921

RESULT 18
US-10-750-623-36115
; Sequence 36115, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
```

```
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36115
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Bovine 19866880565425
US-10-750-623-36115

Query Match          64.0%; Score 16; DB 7; Length 1505;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
Db 898 AAACAGACACATGAGGCTAGG 921

RESULT 19
US-10-995-561-13440/c
; Sequence 13440, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13440
; LENGTH: 60754
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13440

Query Match          64.0%; Score 16; DB 7; Length 60754;
Best Local Similarity 79.2%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAAAAGCAGACAGCGAGCCCTCGG 25
Db 55118 AAAAAGCAGACAGCGGCGACGCTG 55095

RESULT 20
US-10-995-561-13197
; Sequence 13197, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13197
; LENGTH: 96128
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(96128)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13197

Query Match          64.0%; Score 16; DB 7; Length 96128;
```



```
Best Local Similarity 79.2%; Pred. No. 4e+02; Indels 5; Gaps 0;
Matches 19; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 2 AAAAAGCACACGCGAGCCTCGG 25
Db 9739 ACAATGCCACAGTGAGCCTCAG 9762

RESULT 21
US-11-121-086-19
; Sequence 19, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19
; LENGTH: 130733
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-19

Query Match 64.0%; Score 16; DB 8; Length 130733;
Best Local Similarity 79.2%; Pred. No. 4.2e+02; Indels 5; Gaps 0;
Matches 19; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCTCG 24
Db 101635 AAAAAACGACACGAGCTTCG 101658

RESULT 22
US-11-121-086-7
; Sequence 7, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 162085
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-7

Query Match 64.0%; Score 16; DB 8; Length 162085;
Best Local Similarity 79.2%; Pred. No. 4.3e+02; Indels 5; Gaps 0;
Matches 19; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAAACGACACGCGAGCCTCG 24
Db 23042 AAATACACACAGCAGCCTTCG 23065

RESULT 23
US-11-121-086-4/c
; Sequence 4, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
```

```
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4
; LENGTH: 164810
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-4

Query Match 64.0%; Score 16; DB 8; Length 164810;
Best Local Similarity 79.2%; Pred. No. 4.3e+02; Indels 5; Gaps 0;
Matches 19; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 2 AAAAAGCACACGCGAGCCTCGG 25
Db 41267 AAATACGACACGAGCTCCAG 41244

RESULT 24
US-11-121-849-379393
; Sequence 379393, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 379393
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-379393

Query Match 63.2%; Score 15.8; DB 8; Length 25;
Best Local Similarity 89.5%; Pred. No. 1.4e+02; Indels 2; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 3 AAAACGACACAGCGAGGCC 21
Db 5 AATACACACAGCGAGGCC 23

RESULT 25
US-11-136-527-352769/c
; Sequence 352769, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 352769
; LENGTH: 25
; TYPE: DNA
```

; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Probe  
US-11-136-527-352769

Query Match 63.2%; Score 15.8; DB 8; Length 25;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGC 19  
||||| ||||| ||||| |||||  
Db 25 AAAAAAGGACACAGCTAGC 7

## RESULT 26

US-11-136-527-7464/c  
; Sequence 7464, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7464  
; LENGTH: 1400  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-7464

Query Match 63.2%; Score 15.8; DB 8; Length 1400;  
Best Local Similarity 89.5%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGC 19  
||||| ||||| ||||| |||||  
Db 480 AAAAAAGGACACAGCTAGC 462

## RESULT 27

US-10-750-185-54749/c  
; Sequence 54749, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: WM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 54749  
; LENGTH: 1713  
; TYPE: DNA  
; ORGANISM: Bovine 19866880686743  
US-10-750-185-54749

Query Match 63.2%; Score 15.8; DB 7; Length 1713;  
Best Local Similarity 89.5%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGC 19  
||||| ||||| ||||| |||||  
Db 386 AAAAAAGGACACAGCAAGC 368

## RESULT 28

US-10-750-623-54749/c  
; Sequence 54749, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: WM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 54749  
; LENGTH: 1713  
; TYPE: DNA  
; ORGANISM: Bovine 19866880686743  
US-10-750-623-54749

Query Match 63.2%; Score 15.8; DB 7; Length 1713;  
Best Local Similarity 89.5%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGC 19  
||||| ||||| ||||| |||||  
Db 386 AAAAAAGGACACAGCAAGC 368

## RESULT 29

US-11-136-527-3368/c  
; Sequence 3368, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3368  
; LENGTH: 2383  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-3368

Query Match 63.2%; Score 15.8; DB 8; Length 2383;  
Best Local Similarity 89.5%; Pred. No. 2.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGC 19  
||||| ||||| ||||| |||||  
Db 1463 AAAAAAGGACACAGCTAGC 1445

## RESULT 30

US-10-995-561-62101



```
; SEQ ID NO 29381
; LENGTH: 794
; TYPE: DNA
; ORGANISM: Bovine 19866880796529
US-10-750-185-29381

Query Match          62.4%; Score 15.6; DB 7; Length 794;
Best Local Similarity 81.8%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCT 22
   ||||| ||||| ||||| |||||
Db 618 AAAAAACACAAATCAAGCCCT 597

RESULT 35
US-10-750-623-29381/c
; Sequence 29381, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 29381
; LENGTH: 794
; TYPE: DNA
; ORGANISM: Bovine 19866880796529
US-10-750-623-29381

Query Match          62.4%; Score 15.6; DB 7; Length 794;
Best Local Similarity 81.8%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACAGCGAGCCCT 22
   ||||| ||||| ||||| |||||
Db 618 AAAAAACACAAATCAAGCCCT 597

RESULT 36
US-11-041-776-65/c
; Sequence 65, Application US/11041776
; Publication No. US20050272057A1
; GENERAL INFORMATION:
; APPLICANT: ABRAHAMSEN, MITCHELL
; APPLICANT: FREIJE, WADIAH
; TITLE OF INVENTION: SMALL SEGMENTS OF DNA DETERMINE ANIMAL IDENTITY AND
; FILE REFERENCE: 34579-97951
; CURRENT APPLICATION NUMBER: US/11/041,776
; CURRENT FILING DATE: 2005-01-24
; PRIOR APPLICATION NUMBER: 60/538,791
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/539,728
; PRIOR FILING DATE: 2004-01-26
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 65
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Sus scrofa
; FEATURE:

; NAME/KEY: misc feature
; LOCATION: (1) - (870)
; OTHER INFORMATION: nucleotides 121-990 of LEPR
US-11-041-776-65

Query Match          62.4%; Score 15.6; DB 8; Length 870;
Best Local Similarity 81.8%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAACGACACAGCGAGCCCTC 23
   ||||| ||||| ||||| |||||
Db 528 AAAAAATGACACCAAGTCATC 507

RESULT 37
US-10-750-185-59482
; Sequence 59482, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 59482
; LENGTH: 1045
; TYPE: DNA
; ORGANISM: Bovine 19866881274606
US-10-750-185-59482

Query Match          62.4%; Score 15.6; DB 7; Length 1045;
Best Local Similarity 81.8%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AAACGACACAGCGAGCCCTCGG 25
   ||||| ||||| ||||| |||||
Db 153 AGACGACAGGCGAGTCTCGG 174

RESULT 38
US-10-750-623-59482
; Sequence 59482, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 59482
; LENGTH: 1045
; TYPE: DNA
; ORGANISM: Bovine 19866881274606
```

US-10-750-623-59482

Query Match 62.4%; Score 15.6; DB 7; Length 1045;  
Best Local Similarity 81.8%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0

Qy 4 AAACGACACAGCGAGCCTCGG 25  
153 AGACGACAGGGCGAGTCTCTCGG 174

**RESULT 39**

```

US-11-136-527-132
; Sequence 132, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 132
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-132

```

```
Query Match      62.4%; Score 15.6; DB 8; Length 1197;
Best Local Similarity 81.8%; Pred. No. 3.2e+02;
Matches 18: Conservative 0; Mismatches 4; Indels 0
```

Qy 2 AAAAAAGACACAGCGAGCCCTC 23  
||| ||| ||| ||| ||| ||| ||| |||  
pb 307 AAAAAAGGAGACACCGGGCCCTC 328

**RESULT 40**

```

US-10-750-185-44586
; Sequence 44586, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 44586
; LENGTH: 1490
; TYPE: DNA
; ORGANISM: Bovine
; US-10-750-185-44586

```

Query Match 62.4%; Score 15.6; DB 7; Length 1490;  
Best Local Similarity 81.8%; Pred. No. 3.3e+02;  
Matches 18: Conservative 0; Mismatches 4; Indels 0

Qy 1 AAAAAACGACACAGCGAGCCCT 22

D<sub>b</sub> 1397 AAAAATTGACACAGCAAGCTCT 1418

## RESULT 41

```

US-10-750-623-44586
; Sequence 44586, Application US/10750623
; Publication No. US20050287531A1
;
; GENERAL INFORMATION:
;
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
;
; TITLE OF INVENTION: METHODS
;
; FILE REFERENCE: MM1100-1
;
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
;
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
;
; NUMBER OF SEQ ID NOS: 64922
;
; SOFTWARE: PatentIN version 3.1
;
; SEQ ID NO 44586
; LENGTH: 1490
;
; TYPE: DNA
;
; ORGANISM: Bovine
;
US-10-750-623-44586

```

Query Match	62.4%;	Score 15.6;	DB 7;	Length 1490;
Best Local Similarity	81.8%;	Pred. No. 3.3e+02;		
Matches 18; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY  
1 AAAAAACGACACAGCGAGCCCT 22  
|||  
1397 AAAAAATTGACACAGCAAGCTCT 1418  
db

RESULT 42

```

US-10-750-185-52468/C
; Sequence 52468, Application US/10750185
; Publication No. US200502603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52468
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Bovine
; US-10-750-185-52468

```

Query Match	62.4%	Score 15.6;	DB 7;	Length 1700;
Best Local Similarity	81.8%	Pred. No. 3.4e+02;		
Marches	18.	Conservative	0;	Mismatches 4;
Indels	0;			
Gaps	0;			

Qy 1 AAAAAACGACACAGCGAGCCT 22  
||||| | ||||| |||||  
765 AAAAAAGAAAACAGCAAGCCT 744

```
RESULT 43
US-10-750-623-52468/c
; Sequence 52468, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52468
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Bovine 19866881348662
US-10-750-623-52468

Query Match          62.4%; Score 15.6; DB 7; Length 1700;
Best Local Similarity 81.8%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCAGCAGCGCCCT 22
   ||||| | ||||| |||||
Db 765 AAAAAAGCAGCAGCGCCCT 744

RESULT 44
US-10-750-185-35903/c
; Sequence 35903, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 35903
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Bovine 19866881348662
US-10-750-185-35903

Query Match          62.4%; Score 15.6; DB 7; Length 1771;
Best Local Similarity 81.8%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCAGCAGCGCCCT 22
   ||||| | ||||| |||||
Db 361 AAAAAAGCAGCAGCGCCCT 340

RESULT 45
US-10-750-623-35903/c
; Sequence 35903, Application US/10750623
```

;  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 38  
; LENGTH: 151169  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-38

Query Match 62.4%; Score 15.6; DB 8; Length 151169;  
Best Local Similarity 81.8%; Pred. No. 6.5e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCT 22  
DB 17421 AAAAAACGACGCGCCGCGCAT 17442

RESULT 48  
US-10-517-151-105  
; Sequence 105, Application US/10517151  
; Publication No. US20060019252A1  
; GENERAL INFORMATION:  
; APPLICANT: Furukawa, Yuichi  
; APPLICANT: Oncotherapy Science, Inc.  
; APPLICANT: The University of Tokyo  
; TITLE OF INVENTION: Genes and Polypeptides Relating to Hepatocellular or  
; TITLE OF INVENTION: Colorectal Carcinoma  
; FILE REFERENCE: 082379-000400US  
; CURRENT APPLICATION NUMBER: US/10/517,151  
; CURRENT FILING DATE: 2004-12-06  
; PRIOR APPLICATION NUMBER: US 60/386,985  
; PRIOR FILING DATE: 2002-06-06  
; PRIOR APPLICATION NUMBER: WO PCT/JP03/07070  
; PRIOR FILING DATE: 2003-06-04  
; NUMBER OF SEQ ID NOS: 111  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 105  
; LENGTH: 55  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Artificially synthesized target sequence for siRNA  
US-10-517-151-105

Query Match 61.6%; Score 15.4; DB 6; Length 55;  
Best Local Similarity 76.0%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCTCGG 25  
DB 1 AAAAAACGAAACACCGATGACTGGG 25

RESULT 49  
US-10-310-914A-845/c  
; Sequence 845, Application US/10310914A  
; Publication No. US20060003322A1  
; GENERAL INFORMATION:  
; APPLICANT: Benwich, Isaac  
; APPLICANT: Shiler, Kuzat  
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and  
; TITLE OF INVENTION: uses thereof  
; FILE REFERENCE: 06087.0200.CPUS01  
; CURRENT APPLICATION NUMBER: US/10/310,914A  
; CURRENT FILING DATE: 2002-12-06  
; NUMBER OF SEQ ID NOS: 1388402  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 845  
; LENGTH: 121  
; TYPE: RNA  
; ORGANISM: Human

US-10-310-914A-845

Query Match 61.6%; Score 15.4; DB 7; Length 121;  
Best Local Similarity 76.0%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCTCGG 25  
DB 30 AAAAAACACACAGCTAGCCCGCAG 6

RESULT 50  
US-10-995-561-41053/c  
; Sequence 41053, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 41053  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-41053

Query Match 61.6%; Score 15.4; DB 7; Length 201;  
Best Local Similarity 76.0%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAAAAACGACACGCGCGCCCTCGG 25  
DB 180 AAAAAAATCAGCCAGCTCTCGG 156

Search completed: February 3, 2006, 16:20:04  
Job time : 351.111 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-20

Perfect score: 25

Sequence: 1 aaaaactcctgtaggcttccagt 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_br.\*

8: gb\_ro.\*

9: gb\_ats.\*

10: gb\_sy.\*

11: gb\_un.\*

12: gb\_vi.\*

13: gb\_vt.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	74868	9 AF311213	AP311213 Mus muscu
C 2	25	100.0	178391	9 AC154141	AC154141 Mus muscu
C 3	20.8	83.2	128447	9 AC123864	AC123864 Mus muscu
C 4	20.8	83.2	194350	9 AC140328	AC140328 Mus muscu
C 5	20.2	80.8	97011	8 HS997015	HS997015 Human DNA
C 6	20.2	80.8	143254	5 BX649263	BX649263 Zebrafish
C 7	20.2	80.8	157481	8 AC025125	AC025125 Homo sapi
C 8	20.2	80.8	172000	14 AC034194	AC034194 Homo sapi
C 9	20.2	80.8	198136	14 AC090165	AC090165 Homo sapi
C 10	20.2	80.8	279675	14 AC161462	AC161462 Gallus ga
C 11	19.2	76.8	258	6 CQ662163	CQ662163 Sequence
C 12	19.2	76.8	289	6 CQ693741	CQ693741 Sequence
C 13	19.2	76.8	458	10 BV210905	BV210905 GSFTL_66
C 14	19.2	76.8	545	6 CQ923410	CQ923410 Sequence
C 15	19.2	76.8	562	6 CQ225863	CQ225863 Sequence
C 16	19.2	76.8	562	6 CQ300992	CQ300992 Sequence
C 17	19.2	76.8	562	6 CQ338275	CQ338275 Sequence
C 18	19.2	76.8	711	10 BV565426	BV565426 rcr67d10.

19	19.2	76.8	785	10	BV592074	BV592074
C 20	19.2	76.8	2079	9	AB003502	AB003502 Mus muscu
C 21	19.2	76.8	2167	9	BC028325	BC028325 Mus muscu
C 22	19.2	76.8	2359	9	BC031640	BC031640 Mus muscu
C 23	19.2	76.8	2523	8	BC009503	BC009503 Homo sapi
C 24	19.2	76.8	2527	6	AX833607	AX833607 Sequence
C 25	19.2	76.8	2527	8	AK095483	AK095483 Homo sapi
C 26	19.2	76.8	2586	6	CQ714266	CQ714266 Sequence
C 27	19.2	76.8	2587	6	CQ981521	CQ981521 Sequence
C 28	19.2	76.8	2587	6	CS030980	CS030980 Sequence
C 29	19.2	76.8	2587	6	CS036801	CS036801 Sequence
C 30	19.2	76.8	2587	6	CS039932	CS039932 Sequence
C 31	19.2	76.8	2587	6	CS045753	CS045753 Sequence
C 32	19.2	76.8	2587	6	CS086305	CS086305 Sequence
C 33	19.2	76.8	2587	6	HSGST1	X17644 Human GST1-
C 34	19.2	76.8	5569	6	CQ413058	CQ413058 Sequence
C 35	19.2	76.8	5570	6	CQ492962	CQ492962 Sequence
C 36	19.2	76.8	79158	5	AC160831	AC160831 Xenopus t
C 37	19.2	76.8	107172	8	HS737M10	AL031388 Human DNA
C 38	19.2	76.8	115719	14	AC151438	AC151438 Carolinia
C 39	19.2	76.8	116273	14	AP000596	AP000596 Homo sapi
C 40	19.2	76.8	139376	8	HSU95742	U95742 Human chrom
C 41	19.2	76.8	148412	8	AC010654	AC010654 Homo sapi
C 42	19.2	76.8	150525	6	CS086332	CS086332 Sequence
C 43	19.2	76.8	157746	9	AC107677	AC107677 Mus muscu
C 44	19.2	76.8	159613	8	AP007226	AP007226 Homo sapi
C 45	19.2	76.8	160945	8	AP003420	AP003420 Homo sapi
C 46	19.2	76.8	161009	14	AC163997	AC163997 Mus muscu
C 47	19.2	76.8	163795	8	AP000356	AP000356 Homo sapi
C 48	19.2	76.8	163844	8	AC009048	AC009048 Homo sapi
C 49	19.2	76.8	163865	14	AC019081	AC019081 Homo sapi
C 50	19.2	76.8	165948	5	AC147169	AC147169 Xenopus t
C 51	19.2	76.8	166358	14	AC024184	AC024184 Homo sapi
C 52	19.2	76.8	166536	6	CS086331	CS086331 Sequence
C 53	19.2	76.8	168119	14	AC068686	AC068686 Homo sapi
C 54	19.2	76.8	168237	8	AP001262	AP001262 Homo sapi
C 55	19.2	76.8	169776	8	AP001880	AP001880 Homo sapi
C 56	19.2	76.8	169777	14	AC090778	AC090778 Homo sapi
C 57	19.2	76.8	170387	14	AC134388	AC134388 Pan trogl
C 58	19.2	76.8	170682	14	AC073256	AC073256 Homo sapi
C 59	19.2	76.8	170823	14	AC133549	AC133549 Homo sapi
C 60	19.2	76.8	171480	14	AC097333	AC097333 Pan trogl
C 61	19.2	76.8	172827	8	AC007216	AC007216 Homo sapi
C 62	19.2	76.8	173774	8	AL356495	AL356495 Human DNA
C 63	19.2	76.8	176731	5	AC147209	AC147209 Xenopus t
C 64	19.2	76.8	178727	14	AC149878	AC149878 Xenopus t
C 65	19.2	76.8	183271	8	AC104462	AC104462 Homo sapi
C 66	19.2	76.8	185404	14	AC025409	AC025409 Homo sapi
C 67	19.2	76.8	186299	14	AC155555	AC155555 Zea mays
C 68	19.2	76.8	186505	9	AC116394	AC116394 Mus muscu
C 69	19.2	76.8	192788	14	AC145062	AC145062 Pan trogl
C 70	19.2	76.8	194144	5	AC148460	AC148460 Xenopus t
C 71	19.2	76.8	200105	14	AC105308	AC105308 Pan trogl
C 72	19.2	76.8	204316	9	AC087541	AC087541 Mus muscu
C 73	19.2	76.8	216238	9	AL513356	AL513356 Mouse DNA
C 74	19.2	76.8	233387	14	AC162459	AC162459 Mus muscu
C 75	19.2	76.8	237308	14	AC009593	AC009593 Homo sapi
C 76	19.2	76.8	247254	14	AC114844	AC114844 Rattus no
C 77	18.8	75.2	23776	6	AR659661	AR659661 Sequence
C 78	18.8	75.2	57499	14	AC123371_4	Continuation (5 of
C 79	18.8	75.2	91186	14	AC118415_3	Continuation (4 of
C 80	18.8	75.2	97862	14	AC120823_4	Continuation (5 of
C 81	18.8	75.2	105741	14	AC140160	Continuation (4 of
C 82	18.8	75.2	110000	14	AC109746_3	Continuation (3 of
C 83	18.8	75.2	110000	14	AC112426_2	Continuation (2 of
C 84	18.8	75.2	110000	14	AC115558_1	Continuation (1 of
C 85	18.8	75.2	114431	8	BX842590	BX842590 Rhesus DN
C 86	18.8	75.2	114493	14	AC142483	AC142483 Rattus no
C 87	18.8	75.2	114493	14	AC142483	AC142483 Rattus no
C 88	18.8	75.2	123897	14	AC144684	AC144684 Rattus no
C 89	18.8	75.2	134008	8	AC009892	AC009892 Homo sapi
C 90	18.8	75.2	142045	14	AC141650	AC141650 Rattus no
C 91	18.8	75.2	142579	14	AC136110	AC136110 Rattus no

c 92	18.8	75.2	148497	8	AL139117
c 93	18.8	75.2	152134	5	BX571944
c 94	18.8	75.2	152359	14	AC142359
c 95	18.8	75.2	157487	14	AC159033
c 96	18.8	75.2	165467	14	AC159033
c 97	18.8	75.2	177306	8	CR759950
c 98	18.8	75.2	183939	14	AC121650
c 99	18.8	75.2	183939	9	AC121650
c 100	18.8	75.2	194020	9	AC121808
c 101	18.8	75.2	198817	9	AC124562
c 102	18.8	75.2	216124	14	AC098305
c 103	18.8	75.2	219212	14	AC095817
c 104	18.8	75.2	222939	14	AC111598
c 105	18.8	75.2	229001	14	AC120233
c 106	18.8	75.2	229001	14	AC120233
c 107	18.8	75.2	231670	14	AC142363
c 108	18.8	75.2	233829	9	AC140397
c 109	18.8	75.2	237420	14	AC118158
c 110	18.8	75.2	246043	14	AC105890
c 111	18.8	75.2	247758	14	AC117041
c 112	18.8	75.2	251166	14	AC163060
c 113	18.8	75.2	251589	14	AC111900
c 114	18.8	75.2	254584	14	AC110948
c 115	18.8	75.2	254732	14	AC111955
c 116	18.8	75.2	262023	14	AC160648
c 117	18.8	75.2	262023	14	AC120890
c 118	18.8	75.2	265731	14	AC120890
c 119	18.8	75.2	265731	14	AC120890
c 120	18.8	75.2	265833	14	AC125919
c 121	18.8	75.2	270546	14	AC127017
c 122	18.8	75.2	271418	14	AC113678
c 123	18.8	75.2	276158	14	AC129229
c 124	18.8	75.2	279971	14	AC103468
c 125	18.8	75.2	284196	14	AC110447
c 126	18.8	75.2	287138	14	AC114438
c 127	18.8	75.2	291031	14	AC117885
c 128	18.8	75.2	292888	14	AC095400
c 129	18.8	75.2	295460	14	AC114116
c 130	18.8	75.2	316118	14	AC112820
c 131	18.8	75.2	349750	14	AC109776
c 132	18.8	74.4	609	10	BV001701
c 133	18.6	74.4	6043	5	BC070842
c 134	18.6	74.4	27988	8	AC128713
c 135	18.6	74.4	38968	8	AC000075
c 136	18.6	74.4	39982	6	AR308846
c 137	18.6	74.4	39982	6	AR580265
c 138	18.6	74.4	54569	9	BX571685
c 139	18.6	74.4	63753	14	AC090089
c 140	18.6	74.4	87849	2	AC084453
c 141	18.6	74.4	99486	8	AC090082
c 142	18.6	74.4	110000	15	AE016819_06
c 143	18.6	74.4	110000	15	AE016819_07
c 144	18.6	74.4	119346	14	AC164752
c 145	18.6	74.4	130278	8	AC005664
c 146	18.6	74.4	132634	14	AC123394
c 147	18.6	74.4	135245	8	AC073907
c 148	18.6	74.4	142439	14	AC147432
c 149	18.6	74.4	162047	5	BX784036
c 150	18.6	74.4	162269	9	AC000097
	18.6	74.4	163028	9	AC127258
	18.6	74.4	163050	14	AC136176
	18.6	74.4	163865	14	AC019081
	18.6	74.4	165269	8	AC140951

## ALIGNMENTS

RESULT 1	AF311213	74868 bp	DNA	linear	ROD 20-NOV-2002
LOCUS	Mus musculus BAC 171m12 MESDC1 (Mesdc1) and MESDC2 (Mesdc2) genes,				
DEFINITION	complete cds.				
ACCESSION	AF311213				
VERSION	AF311213.1	GI:11245449			
KEYWORDS	Mus musculus (house mouse)				
SOURCE					

ORGANISM	Mus musculus
REFERENCE	Eukaryota: Metazoa; Chordata: Craniata: Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 74868) Wines,M.E., Lee,L., Katari,M.S., Zhang,L., DeRossi,C., Shi,Y., Perkins,S., Feldman,M., McCombie,W.R. and Holdener,B.C.
TITLE	Identification of mesoderm development (mesd) candidate genes by comparative mapping and genome sequence analysis
JOURNAL	Genomics 72 (1), 88-98 (2001)
REFERENCE	2 (bases 1 to 74868) Perkins,S., Feldman,M., McCombie,W.R. and Holdener,B.C.
AUTHORS	Wines,M.E. Dr. Lee,L., Katari,M.S., Zhang,L., DeRossi,C., Shi,Y., Direct Submission
TITLE	Submitted (27-SEP-2000) Biochemistry and Cell Biology, SUNY at Stony Brook, 346 Center for Molecular Medicine, Stony Brook, NY 11794-5215, USA
JOURNAL	
FEATURES	Location/Qualifiers
source	1..74868 /organism="Mus musculus" /mol_type="genomic DNA" /strain="129/SvJ" /db_xref="taxon:10090" /clone="BAC 171m12"
repeat_region	1207..1421 /note="RepeatMasker: B3" /rpt_family="SINE/B2" /rpt_type="dispersed"
repeat_region	complement(1516..1669) /note="RepeatMasker: B1F" /rpt_family="SINE/Alu"
repeat_region	/rpt_type="dispersed" complement(1815..1961) /note="RepeatMasker: B1_MM" /rpt_family="SINE/Alu"
repeat_region	/rpt_type="dispersed" 3013..3100 /note="RepeatMasker: MSTC" /rpt_family="LTR/MaLR"
repeat_region	/rpt_type="dispersed" 3230..3333 /note="RepeatMasker: MT2A" /rpt_family="LTR/MaLR"
repeat_region	/rpt_type="dispersed" 3334..3824 /note="RepeatMasker: RMER19B" /rpt_family="LTR"
repeat_region	/rpt_type="dispersed" complement(3876..4066) /note="RepeatMasker: ORR1C" /rpt_family="LTR/MaLR"
repeat_region	/rpt_type="dispersed" complement(4923..5029) /note="RepeatMasker: L3" /rpt_family="LINE/CR1"
repeat_region	/rpt_type="dispersed" complement(5083..5152) /note="RepeatMasker: L2" /rpt_family="LINE/L2"
repeat_region	/rpt_type="dispersed" complement(6186..6675) /note="RepeatMasker: B4" /rpt_family="SINE/B4"
repeat_region	/rpt_type="dispersed" complement(6817..7147) /note="RepeatMasker: MTE" /rpt_family="LTR/MaLR"
repeat_region	/rpt_type="dispersed" complement(7364..7474) /note="RepeatMasker: ID5" /rpt_family="SINE/ID"
repeat_region	/rpt_type="dispersed"

```
repeat_region 7869..7979
/note="RepeatMasker: B4"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region complement(8175..8252)
/note="RepeatMasker: MIR"
/rpt_family="SINE/MIR"
/rpt_type=dispersed
repeat_region complement(8388..8590)
/note="RepeatMasker: B3"
/rpt_family="SINE/B2"
/rpt_type=dispersed
repeat_region complement(8595..8713)
/note="RepeatMasker: B1F"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region complement(8728..8952)
/note="RepeatMasker: B3A"
/rpt_family="SINE/B2"
/rpt_type=dispersed
repeat_region 10428..10504
/note="RepeatMasker: RMER15"
/rpt_family="LTR"
/rpt_type=dispersed
repeat_region 10681..10812
/note="RepeatMasker: RSINE1"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 11110..11254
/note="RepeatMasker: B1F"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region complement(12197..12336)
/note="RepeatMasker: RSINE1"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 12593..13088
/note="RepeatMasker: L1MA4"
/rpt_family="LINE/L1"
/rpt_type=dispersed
repeat_region 13168..13341
/note="RepeatMasker: B2_Mm1"
/rpt_family="SINE/B2"
/rpt_type=dispersed
repeat_region 13378..13432
/note="RepeatMasker: B2_Mm2"
/rpt_family="SINE/B2"
/rpt_type=dispersed
repeat_region complement(13952..14145)
/note="RepeatMasker: URR1A"
/rpt_family="DNA/MER1_type"
/rpt_type=dispersed
repeat_region 14384..14477
/note="RepeatMasker: MIR"
/rpt_family="SINE/MIR"
/rpt_type=dispersed
repeat_region 15620..15747
/note="RepeatMasker: RSINE1"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 16036..16115
/note="RepeatMasker: MIR"
/rpt_family="SINE/MIR"
/rpt_type=dispersed
repeat_region 16237..16368
/note="RepeatMasker: B1F"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region complement(16622..16763)
/note="RepeatMasker: MIR3"
/rpt_family="SINE/MIR"
/rpt_type=dispersed
repeat_region 17031..17087
```

```
repeat_region /note="RepeatMasker: PB1D10"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region complement(18601..18893)
/note="RepeatMasker: B4A"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 19072..19258
/note="RepeatMasker: B3"
/rpt_family="SINE/B2"
/rpt_type=dispersed
repeat_region complement(19709..19886)
/note="RepeatMasker: B4A"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 20177..20657
/note="RepeatMasker: RMER4"
/rpt_family="LTR"
/rpt_type=dispersed
repeat_region 21857..22102
/note="RepeatMasker: RMER15"
/rpt_family="LTR"
/rpt_type=dispersed
repeat_region 22456..22585
/note="RepeatMasker: B1_MM"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region complement(22971..23063)
/note="RepeatMasker: RMER15"
/rpt_family="LTR"
/rpt_type=dispersed
repeat_region 23529..23653
/note="RepeatMasker: RSINE1"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region 23690..23707
/note="RepeatMasker: RSINE1"
/rpt_family="SINE/B4"
/rpt_type=dispersed
repeat_region complement(23733..23870)
/note="RepeatMasker: B1_MM"
/rpt_family="SINE/Alu"
/rpt_type=dispersed
repeat_region 24067..24243
/note="RepeatMasker: B3"
/rpt_family="SINE/B2"
/rpt_type=dispersed
```

Query Match 100.0%; Score 25; DB 9; Length 74868;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25  
|||||  
Db 33482 AAAAACTCTGTAGGCTTTCAGTG 33458  
|||||

RESULT 2  
AC154141 178391 bp DNA linear ROD 21-JUN-2005  
LOCUS Mus musculus chromosome 7, clone RP24-343G19, complete sequence.  
DEFINITION  
AC154141  
ACCESSION  
AC154141.7 GI:68051863  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 178391)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus chromosome 7, clone RP24-343G19  
JOURNAL Unpublished

REFERENCE  
AUTHORS

2 (bases 1 to 178391)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
 Bloom,T., Boguslavsky,L., Boukhgaltet,B., Camarata,J., Chang,J.,  
 Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
 DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
 Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
 Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
 Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
 MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
 Zimmer,A. and Zody,M.  
 Direct Submission

TITLE  
JOURNAL

Submitted (24-DEC-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

3 (bases 1 to 178391)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
 Bloom,T., Boguslavsky,L., Boukhgaltet,B., Camarata,J., Chang,J.,  
 Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
 DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
 Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
 Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
 Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
 MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
 Zimmer,A. and Zody,M.  
 Direct Submission

TITLE  
JOURNAL

Submitted (28-APR-2005) Broad Institute of MIT and Harvard, 320  
 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

4 (bases 1 to 178391)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
 Bloom,T., Boguslavsky,L., Boukhgaltet,B., Camarata,J., Chang,J.,  
 Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
 DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
 Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
 Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagopian,D., Hages,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
 Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
 MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
 Zimmer,A. and Zody,M.  
 Direct Submission

## JOURNAL

Submitted (21-JUN-2005) Broad Institute of MIT and Harvard, 320  
 Charles Street, Cambridge, MA 02141, USA  
 On Jun 21, 2005 this sequence version replaced gi:62945640.

## COMMENT

All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Broad Institute of MIT and Harvard  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@broad.mit.edu  
 ----- Project Information  
 Center project name: L30897  
 Center clone name: 343\_G\_19  
 -----

## FEATURES

## source

## Location/Qualifiers

1..178391  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="7"  
 /map="7"  
 /clone="RP24-343G19"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 1585..1757  
 /rpt\_family="GA-rich"  
 complement(2018..2295)  
 /rpt\_family="MLTIG1"  
 2324..2371  
 /rpt\_family="GA-rich"  
 2517..2624  
 /rpt\_family="PB1D7"  
 complement(3040..3099)  
 /rpt\_family="MTC"  
 4765..5049  
 /rpt\_family="B4"  
 6950..7018  
 /rpt\_family="CTGGGGn"  
 7195..7229  
 /rpt\_family="TGn"  
 7258..7451  
 /rpt\_family="ID\_B1"  
 9551..9639  
 /rpt\_family="CT-rich"  
 9925..9992  
 /rpt\_family="MIR"  
 complement(10082..10496)  
 /rpt\_family="MTC"  
 10634..10684  
 /rpt\_family="TGn"  
 10684..10745  
 /rpt\_family="GA)n"  
 10747..10819  
 /rpt\_family="TCTG)n"  
 complement(10821..11112)  
 /rpt\_family="B4"  
 12146..12231  
 /rpt\_family="GA-rich"  
 12244..12428  
 /rpt\_family="GGGAGA)n"  
 14033..14105  
 /rpt\_family="TC)n"  
 14105..14170  
 /rpt\_family="TTCC)n"  
 14433..14611  
 /rpt\_family="L2"  
 complement(14959..15592)  
 /rpt\_family="L1MB8"  
 complement(16298..16382)  
 /rpt\_family="Bif"  
 16740..16858  
 /rpt\_family="L1\_MM"  
 16859..16915

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region



```
repeat_region /rpt_family="B4"
1590..1774
/rpt_family="B2"
1809..1863
/rpt_family="B4"
1864..2215
/rpt_family="MaLR"
2216..2362
/rpt_family="Alu"
2363..2394
/rpt_family="MaLR"
2408..2499
/rpt_family="B4"
2446..2517
/rpt_family="tRNA-Ala-GCY_"
2604..2750
/rpt_family="Alu"
3000..3080
/rpt_family="B4"
3081..3197
/rpt_family="Alu"
3270..3425
/rpt_family="B4"
3426..3540
/rpt_family="B4"
3580..3650
/rpt_family="B4"
3698..4076
/rpt_family="MaLR"
4639..4832
/rpt_family="B4"
4911..5013
/rpt_family="B4"
6445..6610
/rpt_family="B2"
6671..6769
/rpt_family="Alu"
7257..7449
/rpt_family="B2"
7545..7736
/rpt_family="B2"
7551..7624
/products="tRNA-Ser"
/notes="Likely pseudogene (HMM Sc=40.07 / Sec struct
Sc=-13.49)"
repeat_region 7784..7955
/rpt_family="B2"
8011..8158
/rpt_family="L1"
8161..8471
/rpt_family="MER1_type"
8588..8833
/rpt_family="B4"
9003..9082
/rpt_family="MER1_type"
9990..10129
/rpt_family="Alu"
9998..10170
/rpt_family="B4"
10451..10585
/rpt_family="Alu"
11281..11356
/rpt_family="ERV1"
12025..12214
/rpt_family="B2"
12217..12344
/rpt_family="Alu"
12498..12636
/rpt_family="Alu"
12681..12835
/rpt_family="B4"
13289..13498
/rpt_family="B2"
```

```
repeat_region 13600..13676
/rpt_family="Alu"
13731..13839
/rpt_family="Alu"
13984..14128
/rpt_family="Alu"
14135..14239
/rpt_family="B4"
14416..14590
/rpt_family="B4"
14622..14801
/rpt_family="B2"
14817..14963
/rpt_family="Alu"
15183..15373
/rpt_family="B2"
15416..15567
/rpt_family="B4"
15565..15663
/rpt_family="B4"
15714..15860
/rpt_family="Alu"
16241..16338
/rpt_family="BC1_MM"
16425..16554
/rpt_family="MER1_type"
16589..16789
/rpt_family="B2"
17586..17724
/rpt_family="B4"
17741..17922
/rpt_family="B2"
17937..17994

Query Match 83.2%; Score 20.8; DB 9; Length 128447;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGCTTCAGT 24
DB 90615 AAAAACTCCTGTAGCTTACAGT 90638

RESULT 4
AC140328/c AC140328 194350 bp DNA linear ROD 27-JAN-2005
LOCUS Mus musculus BAC clone RP23-404L15 from 8, complete sequence.
DEFINITION AC140328
ACCESSION AC140328
VERSION AC140328.3 GI:51854628
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 194350)
AUTHORS Waligorski,J., Kozlowski,A. and Haglund,K.
TITLE The sequence of Mus musculus BAC clone RP23-404L15
JOURNAL Unpublished (2001)
REFERENCE
2 (bases 1 to 194350)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
3 (bases 1 to 194350)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
4 (bases 1 to 194350)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2004) Genome Sequencing Center, 4444 Forest Park
```



# REFERENCE AUTHORS TITLE JOURNAL

Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 194350)  
Wilson, R.K.  
Direct Submission  
Submitted (27-JAN-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Sep 2, 2004 this sequence version replaced gi:50511686.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Summary Statistics  
Center project name: M\_BA0404L15  
-----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Oseguwa  
and Minako Tateno in the laboratory of Pieter de Jong  
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
brain genomic DNA. The clone and detailed information can be  
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

## FEATURES source

1. .194350  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="8"  
/map="8"  
/clone="RP23-404L15"  
/clone\_lib="RPCI-23"

## repeat\_region

257. .346  
/rpt\_family="B4"

## repeat\_region

383. .518  
/rpt\_family="B4"

## repeat\_region

1111. .1252  
/rpt\_family="MIR"

## repeat\_region

1303. .1563  
/rpt\_family="B4"

## repeat\_region

1574. .1710  
/rpt\_family="Alu"

## repeat\_region

1952. .2091  
/rpt\_family="B4"

## repeat\_region

2783. .2866  
/rpt\_family="B4"

## repeat\_region

3025. .3143  
/rpt\_family="Alu"

## repeat\_region

3217. .3320  
/rpt\_family="Alu"

## repeat\_region

3462. .3621  
/rpt\_family="B2"

## repeat\_region

3726. .3862

repeat\_region  
3915. .4120  
/rpt\_family="B2"  
repeat\_region  
4160. .4295  
/rpt\_family="Alu"  
repeat\_region  
4575. .5085  
/rpt\_family="L1"  
repeat\_region  
5359. .5455  
/rpt\_family="B4"  
repeat\_region  
5479. .5519  
/rpt\_family="B4"  
repeat\_region  
5520. .5618  
/rpt\_family="Alu"  
repeat\_region  
5807. .5960  
/rpt\_family="B4"  
repeat\_region  
5983. .6133  
/rpt\_family="Alu"  
repeat\_region  
6144. .6189  
/rpt\_family="L1"  
repeat\_region  
6324. .6437  
/rpt\_family="Alu"  
repeat\_region  
6570. .6709  
/rpt\_family="B2"  
repeat\_region  
6775. .6919  
/rpt\_family="Alu"  
repeat\_region  
6942. .7134  
/rpt\_family="L1"  
repeat\_region  
8118. .8523  
/rpt\_family="MaLR"  
repeat\_region  
8582. .8720  
/rpt\_family="Alu"  
repeat\_region  
8923. .9152  
/rpt\_family="MaLR"  
repeat\_region  
10037. .10170  
/rpt\_family="Alu"  
repeat\_region  
10191. .10337  
/rpt\_family="L2"  
repeat\_region  
11563. .11692  
/rpt\_family="B4"  
repeat\_region  
12166. .12289  
/rpt\_family="Alu"  
repeat\_region  
12748. .12873  
/rpt\_family="Alu"  
repeat\_region  
12874. .12952  
/rpt\_family="ID"  
12878. .12948  
/product="tRNA-Phe"  
/note="Likely pseudogene (HMM Sc=23.61 / Sec struct  
Sc=0.14)"  
repeat\_region  
12953. .12963  
/rpt\_family="Alu"  
repeat\_region  
13038. .13150  
/rpt\_family="Alu"  
repeat\_region  
13882. .13984  
/rpt\_family="Alu"  
repeat\_region  
14084. .14237  
/rpt\_family="L1"  
repeat\_region  
14504. .14907  
/rpt\_family="L1"  
repeat\_region  
14908. .15075  
/rpt\_family="L1"  
repeat\_region  
15084. .15206  
/rpt\_family="Alu"  
repeat\_region  
15223. .15377  
/rpt\_family="B2"  
repeat\_region  
15377. .15471  
/rpt\_family="Alu"  
repeat\_region  
15996. .16132  
/rpt\_family="B4"  
repeat\_region  
16162. .16310  
/rpt\_family="Alu"  
16470. .16520  
/note="Sequence derived from one plasmid subclone."  
unsure

```
repeat_region 16530..16667
/rpt_family="Alu"
repeat_region 16706..16746
/rpt_family="Alu"
repeat_region 16828..17037
/rpt_family="B2"
misc_feature 17033..17063
/note="Sequence derived from PCR product of genomic DNA."
repeat_region 17048..17194
/rpt_family="Alu"
repeat_region 17296..17367
/rpt_family="Alu"
repeat_region 17669..17906
/rpt_family="ERVK"
repeat_region 17907..18035
/rpt_family="Alu"
repeat_region 18054..18351
/rpt_family="B4"
repeat_region 18599..18738
/rpt_family="Alu"
repeat_region 18813..18884
/rpt_family="BC1_MM"
repeat_region 18829..18886
/rpt_family="tRNA-Ala-GCY_"
repeat_region 19023..19150
/rpt_family="B4"
```

```
Query Match 83.2%; Score 20.8; DB 9; Length 194350;
Best Local Similarity 91.7%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTCTGTAGCTTTTCAGT 24
|||||
Db 148168 AAAAACTCTGTAGCTTTACAGT 148145
```

```
RESULT 5
HS997015/c
LOCUS HS997015 97011 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RPS-997015 on chromosome 20p11.21-12
Contains genomic marker D20S904, STSs and GSSs, complete sequence.
ACCESSION AL035073
VERSION AL035073.7 GI:11493139
KEYWORDS HTG; D20S904.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 97011)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 30, 2000 this sequence version replaced gi:9795193.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RPS-997015 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
```

Contact: vegas@sanger.ac.uk

-----  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

#### FEATURES

```
source
location/Qualifiers
1..97011
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="20"
/map="p11.21-12"
/clone="RPS-997015"
/clone_lib="RPCI-5"
misc_feature 1
/note="Clone_left_end: RPS-997015"
misc_feature 97011
/note="Clone_right_end: RPS-997015"
```

#### ORIGIN

```
Query Match 80.8%; Score 20.2; DB 8; Length 97011;
Best Local Similarity 88.0%; Pred. No. 45;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTCTGTAGCTTTTCAGT 25
|||||
Db 54155 AGAAAACTCTGTAGGCATTCACGT 54131
```

#### RESULT 6

EX649263/c

LOCUS

DEFINITION

EX649263

ACCESSION

VERSION

EX649263.6 GI:42415036

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE

1 (bases 1 to 143254)

AUTHORS

Pelan,S.

TITLE

Direct Submission

JOURNAL

Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk

zfish-help@sanger.ac.uk

http://www.sanger.ac.uk/Projects/D\_rerio/fags.shtml#dataeight

On Feb 5, 2004 this sequence version replaced gi:41392404.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

-----

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep

Zebrafish pUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where

this is found the longest good quality representation will be

submitted.

DKEY-21022 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate



http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7767  
Center clone name: 147\_K\_16  
-----

## FEATURES

source	Location/Qualifiers
repeat_region	1..157481 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="RP11-147K16" /clone_lib="RPCI-11 Human Male BAC" complement(49..183) /rpt_family="MIR3"
repeat_region	560..671 /rpt_family="MIR"
repeat_region	complement(1322..1611) /rpt_family="AluSx"
repeat_region	complement(1971..2254) /rpt_family="AluSx"
repeat_region	2257..2321 /rpt_family="MIR"
repeat_region	2359..2484 /rpt_family="MER4C"
repeat_region	complement(2485..2784) /rpt_family="AluSx"
repeat_region	2785..2891 /rpt_family="MER4C"
repeat_region	complement(2934..3246) /rpt_family="AluSx"
repeat_region	3266..3458 /rpt_family="MER4C"
repeat_region	3794..3838 /rpt_family="(CCCCAG)n"
repeat_region	complement(3913..4235) /rpt_family="AluSx"
repeat_region	4645..4951 /rpt_family="AluSg"
repeat_region	complement(5564..5677) /rpt_family="L2"
repeat_region	complement(6590..6691) /rpt_family="L2"
repeat_region	6694..7004 /rpt_family="AluSx"
repeat_region	complement(7564..7866) /rpt_family="AluSx"
repeat_region	8283..8405 /rpt_family="L2"
repeat_region	complement(8851..9002) /rpt_family="AluSc"
repeat_region	9003..9078 /rpt_family="MIR"
repeat_region	9168..9190 /rpt_family="AT rich"
repeat_region	complement(9254..9578) /rpt_family="AluJb"
repeat_region	complement(9735..10065) /rpt_family="L1MC4a"
repeat_region	10066..10358 /rpt_family="AluSx"
repeat_region	complement(10359..10724) /rpt_family="L1MC4a"
repeat_region	complement(10744..11032) /rpt_family="AluY"
repeat_region	complement(11038..11328) /rpt_family="AluSx"
repeat_region	11426..11546 /rpt_family="AluJo"
repeat_region	complement(11556..11684) /rpt_family="L1MB4A"
repeat_region	11691..11855 /rpt_family="AluJo"
repeat_region	complement(12336..12623) /rpt_family="AluSg"
repeat_region	12759..13059 /rpt_family="L2"
repeat_region	13409..13450 /rpt_family="G-rich"
repeat_region	13451..13792 /rpt_family="GA-rich"
unsure	complement(13578..13585) /note="<30 qual SNGL region"
repeat_region	13793..13822 /rpt_family="polypurine"
repeat_region	complement(14075..14128) /rpt_family="MIR"
repeat_region	complement(14533..14679) /rpt_family="AluSg/x"
repeat_region	complement(14943..15153) /rpt_family="L2"
repeat_region	15154..15453 /rpt_family="AluSc"
repeat_region	complement(15454..15856) /rpt_family="L2"
repeat_region	complement(17742..17863) /rpt_family="MER5B"
repeat_region	complement(17864..18176) /rpt_family="AluSg"
repeat_region	complement(18177..18239) /rpt_family="MER5B"
repeat_region	complement(19538..19711) /rpt_family="MIR3"
repeat_region	20980..21075 /rpt_family="MIR"
repeat_region	21095..21219 /rpt_family="(CCCA)n"
repeat_region	21232..21365 /rpt_family="(TCCA)n"
repeat_region	complement(21366..21764) /rpt_family="L2"
repeat_region	22307..22357 /rpt_family="Alu"
repeat_region	complement(22450..22753) /rpt_family="AluSg"
repeat_region	23036..23328 /rpt_family="AluSx"
repeat_region	complement(23329..23607) /rpt_family="L1MB2"
repeat_region	23632..23912 /rpt_family="AluSx"
repeat_region	complement(24128..24260) /rpt_family="FLAM C"
repeat_region	complement(24797..24853) /rpt_family="MIR"
repeat_region	24992..25025 /rpt_family="MER63C"
repeat_region	25026..25056

Query Match 80.8%; Score 20.2; DB 8; Length 157481;  
Best Local Similarity 88.0%; Pred.No. 45;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGTG 25  
|||||  
Db 74363 AAAAGCTCCTGTAGGCTTTGACTG 74387

RESULT 8  
AC024194/c AC024194 172000 bp DNA linear HTG 07-JUL-2000  
LOCUS Homo sapiens chromosome 17 clone RP11-587F22, WORKING DRAFT  
DEFINITION

SEQUENCE, 13 unordered pieces.  
AC024194  
AC024194.2 GI:8099107  
HTG: HTGS PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
REFERENCE  
1 (bases 1 to 172000)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
REFERENCE  
2 (bases 1 to 172000)  
Waterston,R.H.  
Direct Submission  
Submitted (25-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT  
On May 26, 2000 this sequence version replaced gi:7105580.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H NH0587F22  
----- Summary Statistics -----  
Sequencing vector: M13, 100%  
Sequencing vector: plasmid, 0%  
Chemistry: Dye-terminator Big Dye, 0% of reads  
Chemistry: Dye-terminator Big Dye, 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 164472 bases at least Q40  
Consensus quality: 166662 bases at least Q30  
Consensus quality: 167979 bases at least Q20  
Insert size: 186000; agarose-fp  
Insert size: 170800; sum-of-contigs  
Quality coverage: 4.61 in Q20 bases; agarose-fp  
Quality coverage: 5.07 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1310: contig of 1310 bp in length  
\* 1311 1410: gap of unknown length  
\* 1411 2795: contig of 1385 bp in length  
\* 2796 2895: gap of unknown length  
\* 2896 3965: contig of 1071 bp in length  
\* 3967 4066: gap of unknown length  
\* 4067 5681: contig of 1615 bp in length  
\* 5682 5782: gap of unknown length  
\* 5782 7679: contig of 1797 bp in length  
\* 7679 11077: gap of unknown length  
\* 11077 11177: gap of unknown length  
\* 11177 21069: contig of 9892 bp in length  
\* 21069 21169: gap of unknown length  
\* 21169 35301: contig of 14133 bp in length  
\* 35301 35401: gap of unknown length  
\* 35401 48420: contig of 13019 bp in length  
\* 48420 48521: gap of unknown length  
\* 48521 61130: contig of 12610 bp in length  
\* 61130 61231: gap of unknown length  
\* 61231 93885: contig of 32654 bp in length  
\* 93885 93985: gap of unknown length  
\* 93985 125772: contig of 31788 bp in length  
\* 125772 125773: gap of unknown length

FEATURES  
source \* 125873 172000: contig of 46128 bp in length.  
1. .172000  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="17"  
/clone="RP11-587F22"  
1. .1310  
/note="assembly\_name:Contig45"  
1311. .1410  
/estimated\_length=unknown  
1411. .2795  
/note="assembly\_name:Contig48"  
2796. .2895  
/estimated\_length=unknown  
2896. .3966  
/note="assembly\_name:Contig49"  
3967. .4066  
/estimated\_length=unknown  
4067. .5681  
/note="assembly\_name:Contig58"  
5682. .5781  
/estimated\_length=unknown  
5782. .7578  
/note="assembly\_name:Contig59"  
clone end:T7  
vector\_side:right"  
7579. .7678  
/estimated\_length=unknown  
7679. .11076  
/note="assembly\_name:Contig60"  
clone end:SP6  
vector\_side:right"  
11077. .11176  
/estimated\_length=unknown  
11177. .21068  
/note="assembly\_name:Contig61"  
21069. .21168  
/estimated\_length=unknown  
21169. .35301  
/note="assembly\_name:Contig62"  
35302. .35401  
/estimated\_length=unknown  
35402. .48420  
/note="assembly\_name:Contig63"  
48421. .48520  
/estimated\_length=unknown  
48521. .61130  
/note="assembly\_name:Contig64"  
61131. .61230  
/estimated\_length=unknown  
61231. .93884  
/note="assembly\_name:Contig65"  
93885. .93984  
/estimated\_length=unknown  
93985. .125772  
/note="assembly\_name:Contig66"  
125773. .125872  
/estimated\_length=unknown  
125873. .172000  
/note="assembly\_name:Contig67"  
ORIGIN  
Query Match 80.8%; Score 20.2; DB 14; Length 172000;  
Best Local Similarity 88.0%; Pred. No. 45;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAAGCTCTGTAGGCTTTTCACTG 25  
Db 45841 AAAAAAGCTCTGTAGGCTTTTCACTG 45817

```

RESULT 9
AC090165/c
LOCUS
DEFINITION
    Homo sapiens chromosome 17 clone RP11-587F22 map 17, WORKING DRAFT
    SEQUENCE, 32 unordered pieces.
ACCESSION
    AC090165
VERSION
    AC090165.2 GI:14210559
KEYWORDS
    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
    Homo sapiens
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Homidae; Homo.
REFERENCE
    1 (bases 1 to 198136)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
    Barna,N., Bastien,V., Boguslavsky,I., Boukhgalter,B., Brown,A.,
    Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
    Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
    Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
    Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N.,
    Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
    Jones,C., Karatas,A., Lakoque,K., Lamazares,R., Landers,T.,
    Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
    Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
    McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
    Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
    O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
    Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
    Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
    Roy,A., Santos,R., Schauer,S., Schuppback,R., Seaman,S., Severy,P.,
    Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
    Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
    Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
    Zembek,L., Zimmer,A. and Zody,N.
    Direct Submission
    Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    On May 27, 2001 this sequence version replaced gi:12957795.
    All repeats were identified using RepeatMasker:
    Smit, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html
    ----- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIER
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
    ----- Project Information
    Center project name: L12036
    Center clone name: 587_F_22
    ----- Summary Statistics
    Sequencing vector: Plasmid, n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 183444 bases at least Q40
    Consensus quality: 190666 bases at least Q30
    Consensus quality: 193645 bases at least Q20
    Insert size: 163000; agarose-fp
    Insert size: 195036; sum-of-contigs
    Quality coverage: 9.8 in Q20 bases; agarose-fp
    Quality coverage: 8.2 in Q20 bases; sum-of-contigs
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 32 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    
```

```

* be preserved.
* 1
* 49624: contig of 49624 bp in length
* 49625 49724: gap of 100 bp
* 49725 50559: contig of 835 bp in length
* 50559 50659: gap of 100 bp
* 50659 51365: contig of 706 bp in length
* 51365 51465: gap of 100 bp
* 51465 52165: contig of 700 bp in length
* 52165 52265: gap of 100 bp
* 52265 53118: contig of 853 bp in length
* 53118 53219: gap of 100 bp
* 53219 53967: contig of 749 bp in length
* 53967 54067: gap of 100 bp
* 54067 55196: contig of 1129 bp in length
* 55196 55297: gap of 100 bp
* 55297 56063: contig of 767 bp in length
* 56063 56163: gap of 100 bp
* 56163 56986: contig of 823 bp in length
* 56986 57086: gap of 100 bp
* 57086 57758: contig of 672 bp in length
* 57758 57858: gap of 100 bp
* 57858 58532: contig of 674 bp in length
* 58532 58632: gap of 100 bp
* 58632 59985: contig of 1353 bp in length
* 59985 60085: gap of 100 bp
* 60085 61366: contig of 1281 bp in length
* 61366 61466: gap of 100 bp
* 61466 62384: contig of 918 bp in length
* 62384 62484: gap of 100 bp
* 62484 63865: contig of 1381 bp in length
* 63865 63965: gap of 100 bp
* 63965 65265: contig of 1300 bp in length
* 65265 65365: gap of 100 bp
* 65365 66995: contig of 1530 bp in length
* 66995 68299: gap of 100 bp
* 68299 68399: contig of 1304 bp in length
* 68399 68399: gap of 100 bp
* 68399 69389: contig of 990 bp in length
* 69389 69489: gap of 100 bp
* 69489 70667: contig of 1178 bp in length
* 70667 70767: gap of 100 bp
* 70767 72600: contig of 1833 bp in length
* 72600 72700: gap of 100 bp
* 72700 73701: contig of 1001 bp in length
* 73701 73801: gap of 100 bp
* 73801 74876: contig of 1075 bp in length
* 74876 74976: gap of 100 bp
* 74976 76790: contig of 1814 bp in length
* 76790 76890: gap of 100 bp
* 76890 78031: contig of 1141 bp in length
* 78031 78131: gap of 100 bp
* 78131 79413: contig of 1282 bp in length
* 79413 79513: gap of 100 bp
* 79513 83211: contig of 3698 bp in length
* 83211 83311: gap of 100 bp
* 83311 86200: contig of 2889 bp in length
* 86200 86300: gap of 100 bp
* 86300 93593: contig of 7233 bp in length
* 93593 93693: gap of 100 bp
* 93693 115200: contig of 21507 bp in length
* 115200 115200: gap of 100 bp
* 115200 172420: contig of 57120 bp in length
* 172420 172521: gap of 100 bp
* 172521 198136: contig of 25616 bp in length.
    Location/Qualifiers
    1..198136
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="17"
    /map="17"
    /clone_lib="RP11-587F22"
    /clone_lib="RPC1-11 Human Male BAC"

```

FEATURES  
source

KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Gallus gallus (Chicken)
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 279675)
TITLE	The sequence of Gallus gallus clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 279675)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (14-MAY-2005) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 279675)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (18-MAY-2005) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On May 18, 2005 this sequence version replaced gi:65144939.
Center	Genome Center
Center code	WUGSC
Web site	http://genome.wustl.edu
Contact	submissions@wustl.edu
Project Information	Project name: J AAL174G12
Summary Statistics	Sequencing vector: M13; 0% Chemistry: Dye-primer ET; 0% of reads Assembly program: Phrap; version 0.990319 Consensus quality: 268528 bases at least Q40 Consensus quality: 272053 bases at least Q30 Consensus quality: 274289 bases at least Q20
NOTE	This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1	1215: contig of 1215 bp in length
1216	1315: gap of unknown length
1316	2766: contig of 1451 bp in length
2767	2866: gap of unknown length
2867	4846: contig of 1980 bp in length
4847	4946: gap of unknown length
4947	7615: contig of 2669 bp in length
7616	7715: gap of unknown length
7716	10399: contig of 2584 bp in length
10300	10399: gap of unknown length
10400	28927: contig of 18528 bp in length
28928	29027: gap of unknown length
29028	54249: contig of 25222 bp in length
54250	53439: gap of unknown length
54350	73186: contig of 18837 bp in length
73187	73286: gap of unknown length
73287	100255: contig of 26969 bp in length
100256	100355: gap of unknown length
100356	146501: contig of 46146 bp in length
146502	146501: gap of unknown length
146602	214381: contig of 67780 bp in length
214382	214481: gap of unknown length
214482	279675: contig of 65194 bp in length.
Location/Qualifiers	1. 279675
source	
FEATURES	
1..49624	Query Match 80.8%; Score 20.2; DB 14; Length 198136; Best Local Similarity 88.0%; Pred. No. 45; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
/note="assembly_fragment"	
clone end:SP6	
vector side:left"	
49625..49724	
/estimated_length=100	
49725..50559	
/note="assembly_fragment"	
50560..50659	
/estimated_length=100	
50660..51365	
/note="assembly_fragment"	
51366..51465	
/estimated_length=100	
51466..52165	
/note="assembly_fragment"	
52166..52265	
/estimated_length=100	
52266..53118	
/note="assembly_fragment"	
53119..53218	
/estimated_length=100	
53219..53367	
/note="assembly_fragment"	
53368..54067	
/estimated_length=100	
54068..55196	
/note="assembly_fragment"	
55197..55236	
/estimated_length=100	
55237..56063	
/note="assembly_fragment"	
56064..56163	
/estimated_length=100	
56164..56986	
/note="assembly_fragment"	
56987..57086	
/estimated_length=100	
57087..57758	
/note="assembly_fragment"	
57759..57858	
/estimated_length=100	
57859..58532	
/note="assembly_fragment"	
58533..58632	
/estimated_length=100	
58633..59985	
/note="assembly_fragment"	
59986..60085	
/estimated_length=100	
60086..61366	
/note="assembly_fragment"	
61367..61466	
/estimated_length=100	
61467..62384	
/note="assembly_fragment"	
80.8%; Score 20.2; DB 14; Length 198136;	
Best Local Similarity 88.0%; Pred. No. 45;	
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
1 AAAAAGCTCTGTAGGCTTTCAGTG 25	
182201 AAAAAGCTCTGTAGGCTTTCAGTG 182177	
AC161462/c	AC161462
LOCUS	279675 bp
DEFINITION	Gallus gallus chromosome UNKNOWN clone CH261-174G12, WORKING DRAFT
SEQUENCE, 12 unordered pieces.	
AC161462	
AC161462	
GI:66268054	







/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="MAP TO AP000356.1  
EXPRESSED IN ADULT LIVER, SIGNAL = 1.7"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 562;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAATCCTGTAGGCTTTTCAGTG 25  
|||||  
Db 29 AAAAATCCTGTATGCTTTTGATTG 6  
|||||

## RESULT 16

CQ300992/c  
LOCUS CQ300992 562 bp DNA linear PAT 23-JAN-2004  
DEFINITION Sequence 12097 from Patent WO0186003.  
ACCESSION CQ300992  
VERSION CQ300992.1 GI:41261569  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE 1

Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.  
Human genome-derived single exon nucleic acid probes useful for  
analysis of gene expression in human lung  
Patent: WO 0186003-A 12097 15-NOV-2001;  
Aeomica, Inc. (US)

## FEATURES

Location/Qualifiers  
1..562  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="MAP TO AP000356.1  
EXPRESSED IN LUNG, SIGNAL = 1.2"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 562;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAATCCTGTAGGCTTTTCAGTG 25  
|||||  
Db 29 AAAAATCCTGTATGCTTTTGATTG 6  
|||||

## RESULT 17

CQ338275/c  
LOCUS CQ338275 562 bp DNA linear PAT 23-JAN-2004  
DEFINITION Sequence 12369 from Patent WO0157275.  
ACCESSION CQ338275  
VERSION CQ338275.1 GI:41287346  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE 1

Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.  
Human genome-derived single exon nucleic acid probes useful for  
analysis of gene expression in human brain  
Patent: WO 0157275-A 12369 09-AUG-2001;  
Aeomica, Inc. (US)

## FEATURES

Location/Qualifiers  
1..562  
/organism="Homo sapiens"

/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="MAP TO AP000356.1  
EXPRESSED IN BRAIN, SIGNAL = 5.3"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 562;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAATCCTGTAGGCTTTTCAGTG 25  
|||||  
Db 29 AAAAATCCTGTATGCTTTTGATTG 6  
|||||

## RESULT 18

BV565426  
LOCUS BV565426 711 bp DNA linear STS 09-APR-2005  
DEFINITION rc167d10.b1 Clint Pan troglodytes versus STS genomic, sequence  
tagged site.  
ACCESSION BV565426  
VERSION BV565426.1 GI:62456447  
KEYWORDS STS.  
SOURCE Pan troglodytes versus  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.  
1 (bases 1 to 711)  
Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and  
Jaffe, D.B.  
Initial Sequence of the Chimpanzee Genome and Comparison with the  
Human Genome  
Unpublished (2005)

## REFERENCE 1

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 711  
Protocol:  
23,021,928 chimpanzee whole genome shotgun reads were aligned to  
the Human genome NCBI  
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
including Clint (Pan  
troglodytes versus), 3 other Pan troglodytes versus chimps  
(Donald, Karlén, Yvonne), 3 Pan  
troglodytes troglodytes chimps (Noémie, Masuku, Clara) and 2 chimps  
of unknown origin  
(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the  
western chimp and Pan  
troglodytes troglodytes is the central chimp. To be included in  
chimpanzee SNP discovery, a  
read must be at least 500bp in length, at least 50% of its base  
calls must have Phred  
score >= 20, at least 30% of its base calls must satisfy  
SNQS(30,25) (single strand NQS, the  
base in question has Phred score >= 30, the surrounding 10 bases in  
the read have Phred  
score >= 25), and the read must have at least 200 bp SNQS(30,25)  
bases. Reads not uniquely  
placed in the genome and read pairs whose two ends were not  
consistently placed were  
discarded. After above filtering, NQS(30,25) standard was applied  
to all pairs of  
overlapping reads to call NQS bases and SNPs. Alignments (between  
two reads) with less  
than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
exclude alignment between two

copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
1. .711  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1. .>711

## STS

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 711;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAATCTCTGTAGGCTTTCAGTG 25  
Db 315 AAAAATCTCTGTATGCTTTGATTG 338

## RESULT 19

BV592074 785 bp DNA linear STS 12-APR-2005  
LOCUS G591P51714RD5.T0 Clint Pan troglodytes verus STS genomic, sequence tagged site.  
DEFINITION BV592074  
ACCESSION BV592074.1 GI:62507803  
VERSION BV592074  
KEYWORDS STS.

## SOURCE

Pan troglodytes verus  
Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Pan.  
1 (bases 1 to 785)  
Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and Jaffe,D.B.  
Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome  
Unpublished (2005)

## REFERENCE

AUTHORS Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 785  
Protocol:

## JOURNAL

## COMMENT

23, 021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI  
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gen, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQS(30,25)

bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

## source

1. .785  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1. .>785

## STS

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 785;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAATCTCTGTAGGCTTTCAGTG 25  
Db 184 AAAAATCTCTGTATGCTTTGATTG 207

## RESULT 20

## AB003502/c

## LOCUS

## DEFINITION

AB003502 2079 bp mRNA linear ROD 27-AUG-1998  
Mus musculus mRNA for Guanine Nucleotide Regulatory Protein, partial cds.

## ACCESSION

## VERSION

## KEYWORDS

## ORGANISM

AB003502.1 GI:3461879  
Guanine Nucleotide Regulatory Protein.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

## TITLE

1 (bases 1 to 2079)  
Molecular cloning of a novel member of the eukaryotic polypeptide chain-releasing factors (eRF). Its identification as eRF3 interacting with eRF1

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

2 (bases 1 to 2079)  
Direct Submission  
Submitted (02-MAY-1997) Shin-ichi Hoshino, Graduate School of Pharmaceutical Sciences University of Tokyo, Physiological Chemistry; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan  
(E-mail:hoshino@mol.f.u.-tokyo.ac.jp, Tel:03-3812-2111, Fax:03-3815-9604)

## FEATURES

## source

1. .2079  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
<1. .>1653  
/codon\_start=1  
/product="Guanine Nucleotide Regulatory Protein"

## CDS

/protein\_id="BAA32526.1"  
/db\_xref="GI:3461880"  
/translation="VPSFLRGPAPPLSPAGAGDHDGAGGAGGPFVSSQDQSC"

EGSNSTVSMELSEPVNGETEMSPESWEHKEEISEAPPGSGSSGDGRPPRESTOEMM  
EEEREIPPKKANAVAPPGAPKKEHNVNVFIGHVDAGKSTIGGQIMYLTGMVDKRLTKY  
EREAKKRETNWYLSWALDITNOERDKGTVEVGRAYFETEKKHTIILDAPGHKSFPV  
NMIGASQADLAIVISARKGEFETFGKGGQTRHAMLAKTAGVKHLIVILINKWDDP  
TVNWSERYEECKELVPFLKKVGFNPKKDHFMPCSGLTGANLKEQSDPFCFWYIGLP  
FIPYLDNLPNFRNSVDGPRIPLPVDVKYKDMGTVLGKLESGSICKGQQLVMNPKNHY  
EVLGILCDDVETDSVAPGENLKRILKGIIEEELPGFILCDLNNLCHSGRTFDAQIVI  
IHKSLIICPGYNVLIHTICEEVEITALEICLVDPKSGEKSRTFRFRKQDVQVCIARL  
RTAGTICLETFTKDFPQMGFRFLRDEGKTIALGKVLKLVPEKD"

3'UTR polyA\_signal  
1654..2079  
2062..2067

ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 2079;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

|||||  
1 1956 AAAAAATTCCTAGTTTTCATT 1933

RESULT 21  
BC028325/c  
LOCUS  
DEFINITION Mus musculus G1 to S phase transition 1, mRNA (cdna clone MGC:36230 IMAGE:4506483), complete cds.

ACCESSION BC028325  
VERSION BC028325.1 GI:20271481  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 2167)  
Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettner,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 12477932

2 (bases 1 to 2167)  
Director MGC Project.  
Direct Submission  
Submitted (10-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-r@mail.nih.gov](mailto:cgapsb-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E.B. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

REMARK COMMENT

Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.W., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 61 Row: 0 Column: 14.

FEATURES

source

Location/Qualifiers  
1..2167  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:36230 IMAGE:4506483"  
/tissue\_type="Eye, retina, mouse strain C57BL/6"  
/clone\_lib="NIH MGC\_94"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

gene

1..2167  
/gene="Gspt1"  
/note="synonyms: MGC36230, MGC36735"  
/db\_xref="GeneID:14852"  
/db\_xref="MGI:1316728"  
98..1597  
/gene="Gspt1"

CDS

/codon\_start=1  
/product="Gspt1 protein"  
/protein\_id="AAH28325.1"  
/db\_xref="GI:20271482"  
/db\_xref="GeneID:14852"  
/db\_xref="MGI:1316728"  
/translation="MELSEPVNGETEMSPESWEHKEEISEAPPGSGSSGDGRPPE  
ESTQEMEEBEIPKPSAVAPPGAPKKEHNVNVFIGHVDAGKSTIGGQIMYLTGMVD  
KRTLEKREAKKRETNWYLSWALDITNOERDKGTVEVGRAYFETEKKHTIILDAP  
GHKSFPNMIGASQADLAIVISARKGEFETFGKGGQTRHAMLAKTAGVKHLIVL  
INKWDDPTVNWSERYEECKELVPFLKKVGFNPKKDHFMPCSGLTGANLKEQSDPFC  
FWYIGLPFIPYLDNLPNFRNSVDGPRIPLPVDVKYKDMGTVLGKLESGSICKGQQLVM  
NPNHNVETGLSDVETDSVAPGENLKRILKGIIEEELPGFILCDLNNLCHSGRT  
FDAQIVIIHKSLIICPGYNVLIHTICEEVEITALEICLVDPKSGEKSRTFRFRKQDV  
QVCIARLTAGTICLETFTKDFPQMGFRFLRDEGKTIALGKVLKLVPEKD"

ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 2167;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

|||||  
1901 AAAAAATTCCTAGTTTTCATT 1878

RESULT 22

BC031640/c  
LOCUS  
DEFINITION Mus musculus G1 to S phase transition 1, mRNA (cdna clone MGC:36735 IMAGE:4015544), complete cds.

ACCESSION BC031640  
VERSION BC031640.1 GI:21618776  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
1 (bases 1 to 2359)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,



Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 2 Row: h Column: 11

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504166.

```
FEATURES
    source
        1..2523
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="MGC:1735 IMAGE:2822947"
            /tissue_type="lung, small cell carcinoma"
            /clone_lib="NIH MGC 7"
            /lab_host="DH10B-R"
            /note="Vector: pOTB7"
        1..2523
            /gene="GSTP1"
            /note="synonyms: GST1, ETF3A, eRF3a"
            /db_xref="GeneID:2935"
            /db_xref="MIM:139259"
            159..2060
                /gene="GSTP1"
                /codon_start=1
                /product="GSTP1 protein"
                /protein_id="AAH09503.2"
                /db_xref="GI:33874734"
                /db_xref="GeneID:2935"
                /db_xref="MIM:139259"
                /translation="WDPSGGGGGGSSSDSAPDCWDQADMEAPGPGCGGG
                GSLAARAGQRNLAAQSRQLNVAKPVPNVHAAEFVPSFLRCPAAPPPAGGAAN
                NHGAGGAGRAAPVSSQEOESLCEGNSAVSMELSPFENGETEMSPESWEHKEE
                TSEAPFGGSLGDGPPEESAHEMMEEBEIPKPSVAVPAGPKKEHVNVVFIQHVD
                AGKSTIGGQIMLTGVKDRKTLEKREKXNRETWLSWALDTRQBERDKGTVEV
                GRAYPETEKHTFTILDAPGHKS FVPMNIGGASQADLAVLVSARGEFETGFKGGQT
                REHAMLQAGVKHLVLINKWDDPTVANSNRYECKEKLVPFLKKVGFNPKDQIH
                MPCSGLTGNALKEQSDFCFWYGLPFIPLYDLNLPNRSVDGPILPLVDVKYKDMGT
                VLGKUESGICGQGLVMPNKHNVVEVLGILSDDEVDITVAFGENLKLKLGIEEBE
                LFGFTLCDPNNLCGSRGTFDAQIVIEHKHSIICPGYNALVLIHTCIEEVEITALLCLV
                DKKSGEKSTRFRVKVQDQVCILARTAGTICLETFFKFPQPMGRTLRDEGKTIAGK
                VLKLVPEKD"
        774..2048
            /gene="GSTP1"
            /note="TEFI; Region: Translation elongation factor
            EF-1alpha (GTPase) [Translation, ribosomal structure and
            biogenesis]"
            /db_xref="CDD:COG5256"
```

```
ORIGIN
Query Match      76.8%; Score 19.2; DB 8; Length 2523;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGT 24
      ||||| ||||| ||||| ||||| |||||
Db 2375 AAAAAATTCCTGTAGGTTTCATT 2352
```

```
RESULT 24
AX833607/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
    Homo sapiens (human)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE
1
AUTHORS
    Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
```

```
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tanechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
Full-length cDNA sequences
Patent: EP 1347046-A 731 24-SEP-2003;
Research Association for Biotechnology (JP)
Location/Qualifiers
    1..2527
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"

ORIGIN
Query Match      76.8%; Score 19.2; DB 6; Length 2527;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAATCTCTGTAGGCTTTCAGT 24
      ||||| ||||| ||||| ||||| |||||
Db 2405 AAAAAATTCCTGTAGGTTTCATT 2382

RESULT 25
AK095483/c
LOCUS
DEFINITION
    Homo sapiens cDNA FLJ38164 fis, clone DFNES2003742, highly similar
    to G1 TO S PHASE TRANSITION PROTEIN 1.
ACCESSION
AK095483
VERSION
AK095483.1 GI:21754746
KEYWORDS
    oligo capping; fis (full insert sequence).
SOURCE
    Homo sapiens
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Hominidae; Homo.
REFERENCE
1
AUTHORS
    Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
    Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
    Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
    Ishii,S., Yamamoto,J., Saiko,K., Kawai,Y., Isono,Y., Nakamura,Y.,
    Nagahara,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
    Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
    Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
    Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
    Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
    Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
    Hirakawa,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
    Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
    Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
    Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
    Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
    Shiohata,N., Sano,S., Moriya,S., Momiya,H., Satoh,N., Takami,S.,
    Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,T.,
    Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
    Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
    Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
    Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
    Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
    Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
    Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
    Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
    Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
    Nagase,T., Shiraai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
    Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
    Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
    Complete sequencing and characterization of 21,243 full-length
    human cDNAs
    Nat. Genet. 36 (1), 40-45 (2004)
    14702039
REFERENCE
2
JOURNAL
PUBMED
REFERENCE
AUTHORS
    Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
    Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
    Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiya,H.,
```



Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E., Katsoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S., Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Tsuno,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project

## TITLE

## JOURNAL

Unpublished

3 (bases 1 to 2527)

Isogai,T. and Yamamoto,J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

Location/Qualifiers

1..2527

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="DFNES2003742"

/cell\_type="normal dermal fibroblasts (Neonatal Skin)

(NHDP2564)"

/clone\_lib="DFNES2"

/notes="cloning vector: pME18SFL3

primary culture, normal dermal fibroblasts"

ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 2527;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2405 AAAAAATTCCTGTAGGTTTCATT 2382

||||| ||||||| ||||||| ||||||| |||||||

RESULT 26

CQ714266/c

LOCUS

Sequence 200 from Patent WO02069579.

Accession CQ714266

Version CQ714266.1 GI:42275123

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Venter,C.J., Adams M.C., Li,P.W. and Myers,E.W.

Kits, such as nucleic acid arrays, comprising a majority of

humanexons or transcripts, for detecting expression and other uses

thereof

Patent: WO 02069579-A 200 06-SEP-2002;

PE Corporation (NY) (US)

JOURNAL

Patent: WO 02069579-A 200 06-SEP-2002;

Genentech, Inc. (US)

Location/Qualifiers

1..2586

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 27

CQ981521/c

LOCUS

Sequence 376 from Patent EP1498424.

Accession CQ981521

Version CQ981521.1 GI:58190811

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Rosenthal,A., Hermann,K., Heiden,E., Pilarsky,C., Bruemendorf,T., Staub,E., Roepcke,S., Mennerich,D., Kinnemann,H. and Li,X.

Human nucleic acid sequences from lung tumours

Patent: EP 1498424-A 376 19-JAN-2005;

Hinzmann, Bernd (DE); Hermann, Klaus (DE); Heiden, Esmeralda (DE); Rosenthal, Andre (DE)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 28

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 29

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 30

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 31

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 32

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

Location/Qualifiers

1..2587

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 2587;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

||||| ||||||| ||||||| ||||||| |||||||

DB 2463 AAAAAATTCCTGTAGGTTTCATT 2440

||||| ||||||| ||||||| ||||||| |||||||

RESULT 33

CS030980/c

LOCUS

Sequence 486 from Patent WO2005016962.

Accession CS030980

Version CS030980.1 GI:60731250

Keywords Homo sapiens (human)

Source Homo sapiens

Organism

REFERENCE

1

Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.

Compositions and methods for the treatment of immune related

diseases

Patent: WO 2005016962-A 486 24-FEB-2005;

Genentech, Inc. (US)

```
QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 2463 AAAAAATTCCTGTAGGTTTTCATT 2440

RESULT 29
CS036801/c
LOCUS 2587 bp DNA linear PAT 10-MAR-2005
DEFINITION Sequence 6307 from Patent WO2005016962.
ACCESSION CS036801
VERSION CS036801.1 GI:60734216
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Abbas,A., Clark,H., Ouyang,W., Williams,M.P., Wood,W.I. and Wu,T.D.
TITLE Compositions and methods for the treatment of immune related
diseases
JOURNAL Patent: WO 2005016962-A 6307 24-FEB-2005;
Genentech, Inc. (US)
FEATURES
source
1..2587
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 76.8%; Score 19.2; DB 6; Length 2587;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 2463 AAAAAATTCCTGTAGGTTTTCATT 2440

RESULT 30
CS039932/c
LOCUS 2587 bp DNA linear PAT 22-MAR-2005
DEFINITION Sequence 486 from Patent WO2005019258.
ACCESSION CS039932
VERSION CS039932.1 GI:61847833
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Abbas,A., Clark,H., Ouyang,W., Williams,P.M., Wood,W.I. and Wu,T.D.
TITLE Compositions and methods for the treatment of immune related
diseases
JOURNAL Patent: WO 2005019258-A 486 03-MAR-2005;
Genentech, Inc. (US)
FEATURES
source
1..2587
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 76.8%; Score 19.2; DB 6; Length 2587;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 2463 AAAAAATTCCTGTAGGTTTTCATT 2440

RESULT 31
CS045753/c
LOCUS 2587 bp DNA linear PAT 22-MAR-2005
DEFINITION Sequence 6307 from Patent WO2005019258.
ACCESSION CS045753
VERSION CS045753.1 GI:61852019
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Abbas,A., Clark,H., Ouyang,W., Williams,P.M., Wood,W.I. and Wu,T.D.
TITLE Compositions and methods for the treatment of immune related
diseases
JOURNAL Patent: WO 2005019258-A 6307 03-MAR-2005;
Genentech, Inc. (US)
FEATURES
source
1..2587
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 76.8%; Score 19.2; DB 6; Length 2587;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 2463 AAAAAATTCCTGTAGGTTTTCATT 2440

RESULT 32
CS086305/c
LOCUS 2587 bp DNA linear PAT 25-MAY-2005
DEFINITION Sequence 8 from Patent WO2005042786.
ACCESSION CS086305
VERSION CS086305.1 GI:66711877
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Harris,C. and Davis,L.
TITLE Compositions and methods for glioma classification
JOURNAL Patent: WO 2005042786-A 8 12-MAY-2005;
Exagen Diagnostics, Inc. (US)
FEATURES
source
1..2587
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 76.8%; Score 19.2; DB 6; Length 2587;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 2463 AAAAAATTCCTGTAGGTTTTCATT 2440

RESULT 33
HSGST1/c
LOCUS 2587 bp mRNA linear PRI 18-APR-2005
DEFINITION Human GST1-Hs mRNA for GTP-binding protein.
ACCESSION X17644
VERSION X17644.1 GI:31920
```



```
REFERENCE 2 (bases 1 to 79158)
AUTHORS Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2005) Multimegabase Sequencing Center, Institute
for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA
REFERENCE 3 (bases 1 to 79158)
AUTHORS Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2005) Multimegabase Sequencing Center, Institute
for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA
REFERENCE 4 (bases 1 to 79158)
AUTHORS Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2005) Multimegabase Sequencing Center, Institute
for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA
COMMENT On Jul 1, 2005 this sequence version replaced gi:63003944.
FEATURES
    source
        1..79158
            /organism="Xenopus tropicalis"
            /mol_type="genomic DNA"
            /db_xref="taxon:8364"
            /clone="ISB1-45G24"
            1..59739
                /gene="MYT1"
                /note="3' end of MYT1, based on similarity to X. laevis
                cDNA U67078"
            28740..29070
                /note="low quality region"
            30720..31010
                /note="low quality region"
            33360..34730
                /note="low quality region; possible misassembly in VNTR
                repeat"
            40270..40370
                /note="low quality region"
            54510..54590
                /note="low quality region"
            57300..57400
                /note="low quality region"
            57710..57970
                /note="low quality region"
            58700..59090
                /note="low quality region"
            61570..61650
                /note="low quality region"
            63050..63160
                /note="low quality region"
            65060..65500
                /note="low quality region"
            67400..67480
                /note="low quality region"
            69460..69590
                /note="low quality region"
            /note="low quality region"
    ORIGIN
        Query Match 76.8%; Score 19.2; DB 5; Length 79158;
        Best Local Similarity 87.5%; Pred. No. 1.5e+02;
        Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
        QY 1 AAAAAGCTCTGTAGGCTTTTCAGT 24
            |||||
            3707 AACATCTCTGTAGGCTTCCAGT 3684
        Db
    RESULT 37
    LOCUS HS737M10 107172 bp DNA linear PRI 18-MAY-2005
    DEFINITION Human DNA sequence from clone RP4-737M10 on chromosome Xq23
    Contains the 5' end of the TRPC5 gene for transient receptor
    potential cation channel subfamily C member 5, complete sequence.
    AL031388
    ACCESSION AL031388.1 GI:3980442
    VERSION
    KEYWORDS HTG; TRPC5.
```

```
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 107172)
AUTHORS Pearce,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT On Dec 8, 1998 this sequence version replaced gi:3646111.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP4-737M10 is from the library RPCI-4 constructed by the group of
Pietter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
    source
        1..107172
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="RZPD:RPCIP704M10737"
            /db_xref="taxon:9606"
            /map="q23"
            /clone="RP4-737M10"
            /clone_lib="RPCI-4"
            join(complement(60101..60998),
            complement(AC005191.1:116058..116456),
            complement(AC005191.1:76306..76827),
            complement(AC005191.1:17785..18121),
            complement(AC005191.1:16313..16452),
            complement(AC005191.1:11129..11451),
            complement(AL049563.4:111864..112059),
            complement(AL049563.4:58878..59081),
            complement(AL049563.4:58108..58149),
            complement(AL049563.4:55928..56017),
            complement(AL049563.4:51258..53945))
            /gene="TRPC5"
            /locus_tag="RP1-68D15.2-001"
            join(complement(60101..60998),
            complement(AC005191.1:116058..116456),
            complement(AC005191.1:76306..76827),
            complement(AC005191.1:17785..18121),
            complement(AC005191.1:16313..16452),
            complement(AC005191.1:11129..11451),
            complement(AL049563.4:111864..112059),
            complement(AL049563.4:58878..59081),
            complement(AL049563.4:58108..58149),
            complement(AL049563.4:55928..56017),
            complement(AL049563.4:51258..53945))
            /gene="TRPC5"
            /locus_tag="RP1-68D15.2-001"
```

/product="transient receptor potential cation channel, subfamily C, member 5" /note="match: CDNAS: AF054568.1"

ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 107172; Best Local Similarity 87.5%; Pred. No. 1.5e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGCTAGGCTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 68000 AAAAAAATCCTTTAGGCTTTCAT 68023

RESULT 38  
AC151438  
LOCUS  
DEFINITION  
Carollia perspicillata clone 438K12, WORKING DRAFT SEQUENCE, 4 unordered pieces.  
AC151438.1 GI:51784082  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
Carollia perspicillata  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera; Phyllostomidae; Carolliinae; Carollia.  
REFERENCE  
AUTHORS  
Antonellis,A., Ayala,K., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B., Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Pagnuirian,C., Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2. (bases 1 to 115719)  
Green,E.D.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Center: NTH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: iup  
Center clone name: 438K12  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 114321 bases at least Q40  
Consensus quality: 114441 bases at least Q30  
Consensus quality: 114518 bases at least Q20  
Insert size: 183000; agarose-1p  
Insert size: 115419; sum-of-contigs  
Quality coverage: 9.63x in Q20 bases; agarose-1p  
Quality coverage: 15.27x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

\* be preserved.  
\* 1 2618: contig of 2618 bp in length  
\* 2619: gap of unknown length  
\* 2719: contig of 4991 bp in length  
\* 7710: gap of unknown length  
\* 7810: gap of unknown length  
\* 42679: contig of 34870 bp in length  
\* 42680: gap of unknown length  
\* 115719: contig of 72940 bp in length.  
Location/Qualifiers  
1..115719  
/organism="Carollia perspicillata"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:40233"  
/clone\_lib="The Green Lab"  
/clone="438K12"  
/note="egreen@nhgri.nih.gov"  
1..2618  
/note="assembly\_fragment"  
2619..2718  
/estimated\_length=unknown  
2719..7709  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right  
7710..7809  
/estimated\_length=unknown  
7810..42679  
/note="assembly\_fragment"  
42680..42779  
/estimated\_length=unknown  
42780..115719  
/note="assembly\_fragment"  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 115719; Best Local Similarity 87.5%; Pred. No. 1.5e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACTCTGCTAGGCTTTCAGT 25  
||||| ||||| ||||| ||||| |||||  
DB 87533 AAAAAATCCTGTGGGTTTCAGT 87556

RESULT 39  
AP000596/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 11 clone XHPac-662A14 map 11q13, WORKING DRAFT SEQUENCE, 29 unordered pieces.  
ACCESSION  
AP000596.2 GI:8118802  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 116273)  
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
Homo sapiens 116,273 genomic DNA of 11q13  
TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 116273)  
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
Direct Submission  
Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp)  
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924  
On May 31, 2000 this sequence version replaced gi:6997486.  
----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: [hattori@gs.riken.go.jp](mailto:hattori@gs.riken.go.jp)

----- Project Information

Center project name: HumDraft11

Center clone name: XHPac-662A14

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 103806 bases at least Q40

Consensus quality: 109449 bases at least Q30

Consensus quality: 111869 bases at least Q20

Insert size: 113473; sum-of-contigs

Quality coverage: 4.03x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 38138 contig of 38138 bp in length
38239 44698 contig of 4460 bp in length
44799 49514 contig of 4716 bp in length
49615 55534 contig of 5920 bp in length
55635 60136 contig of 5307 bp in length
60237 65623 contig of 4379 bp in length
65724 70102 contig of 1185 bp in length
70203 71387 contig of 1185 bp in length
71488 73839 contig of 2352 bp in length
73940 77366 contig of 3427 bp in length
77467 80065 contig of 2599 bp in length
80166 83825 contig of 3660 bp in length
83926 86777 contig of 2852 bp in length
86878 89783 contig of 2906 bp in length
89884 92751 contig of 2868 bp in length
92852 95068 contig of 1571 bp in length
95169 96739 contig of 1073 bp in length
96840 97912 contig of 100 bp
98013 99744 contig of 1732 bp in length
99845 101153 contig of 1309 bp in length
101254 102868 contig of 1615 bp in length
102969 104921 contig of 100 bp
104922 105021 contig of 1953 bp in length
105022 106500 contig of 1479 bp in length
106501 106600 contig of 100 bp
106601 108529 contig of 1929 bp in length
108530 108629 contig of 100 bp
108630 110841 contig of 2212 bp in length
110842 110941 contig of 100 bp
110942 112367 contig of 1426 bp in length
112368 112468 contig of 100 bp
112469 113665 contig of 1198 bp in length
113666 114974 contig of 1209 bp in length
114975 115074 contig of 100 bp
115075 116273 contig of 1199 bp in length
Sequence updated (26-May-2000).
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 38138 contig of 38138 bp in length
38239 44698 contig of 4460 bp
44799 49514 contig of 4716 bp in length
49615 55534 contig of 5920 bp in length
55635 60136 contig of 5307 bp in length
60237 65623 contig of 4379 bp in length
65724 70102 contig of 1185 bp in length
70203 71387 contig of 1185 bp in length
71488 73839 contig of 2352 bp in length
73940 77366 contig of 3427 bp in length
```

```
* 65624 65723: gap of 100 bp
65724 70102: contig of 4379 bp in length
70103 71387: contig of 1185 bp in length
71388 73839: contig of 2352 bp in length
73840 77366: contig of 3427 bp in length
77367 80065: contig of 2599 bp in length
80066 83825: contig of 3660 bp in length
83826 86777: contig of 2852 bp in length
86778 89783: contig of 2906 bp in length
89784 92751: contig of 2868 bp in length
92752 95068: contig of 1571 bp in length
95069 96739: contig of 1073 bp in length
96740 97912: contig of 100 bp
97913 99744: contig of 1732 bp in length
99745 101153: contig of 1309 bp in length
101154 102868: contig of 1615 bp in length
102869 104921: contig of 100 bp
104922 105021: contig of 1953 bp in length
105022 106500: contig of 1479 bp in length
106501 106600: gap of 100 bp
106601 108529: contig of 1929 bp in length
108530 108629: gap of 100 bp
108630 110841: contig of 2212 bp in length
110842 110941: gap of 100 bp
110942 112367: contig of 1426 bp in length
112368 112468: gap of 100 bp
112469 113665: contig of 1198 bp in length
113666 114974: contig of 1209 bp in length
114975 115074: gap of 100 bp
115075 116273: contig of 1199 bp in length.
```

FEATURES

```
source
1..116273
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q13"
/clone="XHPac-662A14"

misc_feature
1..38138
/note="assembly_fragment"
38239..44698
/note="assembly_fragment"
44799..49514
/note="assembly_fragment"
49615..55534
/note="assembly_fragment"
55635..60136
/note="assembly_fragment"
60237..65623
/note="assembly_fragment"
65724..70102
/note="assembly_fragment"
70203..71387
/note="assembly_fragment"
71488..73839
/note="assembly_fragment"
73940..77366
```

```

misc_feature /note="assembly_fragment"
77467..80065
misc_feature /note="assembly_fragment"
80166..83825
misc_feature /note="assembly_fragment"
83926..86777
misc_feature /note="assembly_fragment"
86878..89783
misc_feature /note="assembly_fragment"
89884..92751
misc_feature /note="assembly_fragment"
92852..95068
misc_feature /note="assembly_fragment"
95169..96739
misc_feature /note="assembly_fragment"
96840..97912
misc_feature /note="assembly_fragment"
98013..99744
misc_feature /note="assembly_fragment"
99845..101153
misc_feature /note="assembly_fragment"
101254..102868
misc_feature /note="assembly_fragment"
102969..104921
misc_feature /note="assembly_fragment"
105022..106500

Query Match 76.8%; Score 19.2; DB 14; Length 116273;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGATGCTTCAGTG 25
DB 99401 AAAAATCTCAGGAGGCTCAGTG 99378

RESULT 40
HSU95742/c
LOCUS HSU95742 139376 bp DNA linear PRI 10-JAN-2000
DEFINITION Human chromosome 16p13.1 BAC clone CIT987SK-551G9 complete
sequence.
ACCESSION U95742
VERSION U95742.1 GI:2339843
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 139376)
Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
Genome duplications and other features in 12 Mb of DNA sequence
from human chromosome 16p and 16q
Genomics 60 (3), 295-308 (1999)
JOURNAL 10493829
PUBMED
REFERENCE 2 (bases 1 to 139376)
Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.,
Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
Human chromosome 16p13.1 BAC clone CIT987SK-551G9 complete sequence
Unpublished
JOURNAL 3 (bases 1 to 139376)
REFERENCE 3 (bases 1 to 139376)
Adams,M.D.
Direct Submission
JOURNAL 28-MAR-1997
TITLE Submitted (28-MAR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 139376)
Adams,M.D., Loftus,B.J., Zhou,L., La Bombard,M., Kim,U.J. and
Venter,J.C.
Direct Submission
JOURNAL 21-AUG-1997
TITLE Submitted (21-AUG-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

COMMENT
Medical Center Dr., Rockville, MD 20850, USA
On Aug 21, 1997 this sequence version replaced gi:1930144.
BAC clone CIT987SK-551G9 is located in band 16p13.1 of chromosome
16. Genes were identified by a combination of five methods: XGRAIL
(available by anonymous ftp from arthur.epm.ornl.gov), Genefinder
(available by anonymous ftp from colin@u.washington.edu), GENSCAN
searches of the EST database at TIGR
(http://www.tigr.org/tdb/hcd/hcd.html) and searches against a
peptide database. Repeats were identified using Censor (Jurka, J.,
Klonowski, P., Dagman, V., Pelton, P. Censor-a program for the
identification and elimination of repetitive elements from DNA
sequences. Computers Chem 20: 119-121 (1996); available by
anonymous ftp from ncbi.nlm.nih.gov).

FEATURES
Source
1..139376
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.1"
/clone="551G9"
complement(849..1113)
/rpt_family="Alu-Sz"
repeat_region
complement(1168..1455)
/rpt_family="Alu-Sg"
repeat_region
1870..2157
/rpt_family="Alu-Sxz"
repeat_region
complement(3106..3396)
/rpt_family="Alu-Sx"
repeat_region
3563..3735
/rpt_family="MIR"
repeat_region
complement(4063..4355)
/rpt_family="Alu-Sx"
repeat_region
4694..4925
/rpt_family="L1MB3"
repeat_region
complement(4933..5218)
/rpt_family="Alu-Jo"
repeat_region
5380..5677
/rpt_family="Alu-Sz"
repeat_region
7958..8247
/rpt_family="Alu-Sz"
repeat_region
8531..8820
/rpt_family="Alu-Jo"
repeat_region
complement(8930..9219)
/rpt_family="Alu-Sx"
repeat_region
9542..9817
/rpt_family="Alu-Sg"
repeat_region
9861..10154
/rpt_family="Alu-Jo"
repeat_region
11127..11388
/rpt_family="Alu-Sx"
repeat_region
12278..12567
/rpt_family="Alu-Jo"
repeat_region
complement(12856..13166)
/rpt_family="Alu-Jo"
repeat_region
13866..13994
/rpt_family="Alu-Jo"
repeat_region
14003..14289
/rpt_family="Alu-Sx"
repeat_region
14505..14791
/rpt_family="Alu-Y"
repeat_region
complement(15165..15307)
/rpt_family="Alu-Jo"
repeat_region
15829..15960
/rpt_family="Alu-Spgxz"
repeat_region
15961..16259
/rpt_family="Alu-Sz"
repeat_region
16262..16333
/rpt_family="L1"
repeat_region
16336..16451
/rpt_family="L1MB3"
repeat_region
16452..16724

```



repeat_region	/rpt_family="LIME3A"	16738..17162	/rpt_family="LIME3A"	16738..17162	/protein_id="AAB67251.1"	
repeat_region	/rpt_family="LMB7"	17288..17346	/rpt_family="LMB7"	17288..17346	/db_xref="GI:1930146"	
repeat_region	/rpt_family="Alu-Sxz"	17405..17633	/rpt_family="Alu-Sxz"	17405..17633	/translation="MLQMAQSQNEYFDSLHACIPQCLRCSSNTPPLTCQRYCNARV VTSVGTGTAIIWTCGLGLIIISLAVFLMFLIKISSEPLKDFKNTGSLGGMANI DLEKSTGDEIILPRGLEVTVECTCEDCIKSKPKVSDHCFPLPAMEGATILVTTK TNDYCKSLPALSAATEIEKSIAR"	
repeat_region	/rpt_family="Alu-Sxz"	complement(17952..18239)	/rpt_family="Alu-Sxz"	complement(17952..18239)	complement(29760..29890)	
repeat_region	/rpt_family="Alu-Sxz"	complement(18302..18447)	/rpt_family="Alu-Sxz"	complement(18302..18447)	29929..30218	
repeat_region	/rpt_family="MIR"	complement(18731..18923)	/rpt_family="MIR"	complement(18731..18923)	/rpt_family="Alu-Jo"	
repeat_region	/rpt_family="Alu-Jo"	21006..21086	/rpt_family="Alu-Jo"	21006..21086	/rpt_family="Alu-Sz"	
repeat_region	/rpt_family="Alu-Sxz"	21182..21233	/rpt_family="Alu-Sxz"	21182..21233	complement(31107..31400)	
repeat_region	/rpt_family="Alu-Jo"	21235..21285	/rpt_family="Alu-Jo"	21235..21285	/rpt_family="Alu-Sx"	
repeat_region	/rpt_family="Alu-Sx"	21341..21633	/rpt_family="Alu-Sx"	21341..21633	32362..32509	
repeat_region	/rpt_family="Alu-Ya5"	21788..22076	/rpt_family="Alu-Ya5"	21788..22076	/rpt_family="Alu-Jb"	
repeat_region	/rpt_family="Alu-Sxz"	complement(22153..22442)	/rpt_family="Alu-Sxz"	complement(22153..22442)	/rpt_family="Alu-Sz"	
repeat_region	/rpt_family="Alu-Sbcg"	complement(22711..22989)	/rpt_family="Alu-Sbcg"	complement(22711..22989)	complement(33184..33469)	
repeat_region	/rpt_family="Alu-Sx"	23440..23656	/rpt_family="Alu-Sx"	23440..23656	/rpt_family="Alu-Jb"	
repeat_region	/rpt_family="Alu-Sq"	23879..24033	/rpt_family="Alu-Sq"	23879..24033	complement(33836..34124)	
repeat_region	/rpt_family="MER4B"	complement(24063..24355)	/rpt_family="MER4B"	complement(24063..24355)		
repeat_region	/rpt_family="Alu-Sp"	complement(24364..24636)	/rpt_family="Alu-Sp"	complement(24364..24636)		
repeat_region	/rpt_family="Alu-Sx"	complement(24651..24940)	/rpt_family="Alu-Sx"	complement(24651..24940)		
repeat_region	/rpt_family="MER4B"	24958..25058	/rpt_family="MER4B"	24958..25058		
repeat_region	/rpt_family="Alu-Sg"	complement(25080..25368)	/rpt_family="Alu-Sg"	complement(25080..25368)		
repeat_region	/rpt_family="MER4B"	25379..25437	/rpt_family="MER4B"	25379..25437		
repeat_region	/rpt_family="Alu-Sz"	25438..25727	/rpt_family="Alu-Sz"	25438..25727		
repeat_region	/rpt_family="MER4A"	25744..25957	/rpt_family="MER4A"	25744..25957		
repeat_region	/rpt_family="Alu-Jb"	26078..26369	/rpt_family="Alu-Jb"	26078..26369		
repeat_region	/rpt_family="Alu-Sg"	26401..26689	/rpt_family="Alu-Sg"	26401..26689		
repeat_region	/rpt_family="Alu-Sx"	26723..27011	/rpt_family="Alu-Sx"	26723..27011		
repeat_region	/rpt_family="Alu-Sx"	complement(27053..27167)	/rpt_family="Alu-Sx"	complement(27053..27167)		
repeat_region	/rpt_family="Alu-Sg"	27205..27359	/rpt_family="Alu-Sg"	27205..27359		
repeat_region	/rpt_family="MIR"	27362..27438	/rpt_family="MIR"	27362..27438		
repeat_region	/rpt_family="Alu-Sg"	27675..27957	/rpt_family="Alu-Sg"	27675..27957		
repeat_region	/rpt_family="Alu-Sxz"	28063..28352	/rpt_family="Alu-Sxz"	28063..28352		
repeat_region	/rpt_family="Alu-Sx"	complement(28745..29034)	/rpt_family="Alu-Sx"	complement(28745..29034)		
gene	/rpt_family="Alu-Sx"	complement(29162..32118)	/rpt_family="Alu-Sx"	complement(29162..32118)		
mRNA	/gene="551G9.1"	complement(join(29162..29659,30884..31030,31771..32118))	/gene="551G9.1"	complement(join(29162..29659,30884..31030,31771..32118))		
CDS	/gene="551G9.1"	complement(join(29382..29659,30884..31030,31771..31900))	/gene="551G9.1"	complement(join(29382..29659,30884..31030,31771..31900))		
	/codon_start=1		/codon_start=1			
	/product="B-cell maturation protein"		/product="B-cell maturation protein"			



Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)  
----- Project Information  
Center project name: LI7624  
Center clone name: 235\_F\_22  
-----

FEATURES

Source

Location/Qualifiers  
1..157746

/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/map="3"  
/clone="RP23-235P22"  
/clone\_lib="RPCI-23 Female Mouse BAC"

complement(243..343)  
/rpt\_family="PB1D7"  
1830..1897

/rpt\_family="TA)n"  
complement(2283..2453)

/rpt\_family="B2\_Mm1"  
complement(2509..2772)

/rpt\_family="L1\_MM"  
2772..3096

/rpt\_family="Lx"  
complement(3734..3904)

/rpt\_family="URR1A"  
3921..4115

/rpt\_family="B4A"  
4154..4223

/rpt\_family="CATA)n"  
5473..5640

/rpt\_family="B3A"  
6264..6445

/rpt\_family="TATG)n"  
complement(6449..6718)

/rpt\_family="B4"  
7602..7737

/rpt\_family="B1\_MM"  
7738..7776

/rpt\_family="CAAAA)n"  
8885..8935

/rpt\_family="AT\_Rich"  
complement(8936..9161)

/rpt\_family="Lx2"  
9164..10365

/rpt\_family="Lx2"  
10367..10403

/rpt\_family="L1\_Cp"  
11094..11172

/rpt\_family="CAAAA)n"  
11872..11909

/rpt\_family="TG)n"  
complement(11913..12056)

/rpt\_family="RSINE1"  
complement(12090..12475)

/rpt\_family="LIM2"  
12491..12528

/rpt\_family="TCC)n"  
complement(12720..13181)

/rpt\_family="LIM2"  
13609..13760

/rpt\_family="CACCC)n"  
15181..15513

/rpt\_family="RLTRETN\_MM"  
15986..16050

/rpt\_family="AT\_Rich"  
16719..16873

/rpt\_family="Zaphod"  
18776..18806

/rpt\_family="TAAA)n"  
19190..19253

repeat\_region  
19718..20155  
/rpt\_family="Lx"  
21103..21138  
/rpt\_family="(TTTG)n"  
21471..21582  
/rpt\_family="MIR3"  
complement(21875..22073)  
/rpt\_family="B3"  
22128..22152  
/rpt\_family="AT\_Rich"  
23624..23668  
/rpt\_family="(TAAA)n"  
24248..24346  
/rpt\_family="Lx9"  
complement(25111..25342)  
/rpt\_family="RLTR11A"  
25518..25794  
/rpt\_family="L1\_MM"  
26204..28475  
/rpt\_family="Lx4"  
28496..28573  
/rpt\_family="(GA)n"  
28574..29813  
/rpt\_family="Lx4"  
complement(30658..30949)  
/rpt\_family="B4"  
complement(31119..31258)  
/rpt\_family="B4"  
complement(31406..31730)  
/rpt\_family="MLT1C"

Query Match 76.8%; Score 19.2; DB 9; Length 157746;  
Best Local Similarity 87.5%; Pred.No.1.5e-02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACTCTGTAGGCTTTTCAGTG 25  
|||||

Db 111564 AAAGACTCTGTAGGCTGTAGTG 111587  
|||||

RESULT 44

AC007226/c

LOCUS

DEFINITION

AC007226

VERSION

AC007226.3

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

AC007226 159613 bp DNA linear PRI 02-DEC-2000  
Homo sapiens chromosome 16 clone RPCI-11\_520B23, complete sequence.

AC007226

HTG.

KEYWORDS

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 159613)

Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,

Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,

Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,

Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,

Han, C. and Deaven, L.

Sequencing of Human Chromosome 16q12

Unpublished

2 (bases 1 to 159613)

Ricke, D.O.

Large Scale Sequence Analysis and Annotation with the Sequence

Comparison Analysis (SCAN) System

Unpublished

3 (bases 1 to 159613)

Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,

Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,

Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,

Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,

McMurry, K., Han, C. and Deaven, L.

Direct Submission



```

/rept_family="Alu"
36844..37133
/rpt_family="Alu"
37265..37529
/rpt_family="Alu"
complement(38294)
/rpt_family="LTR"
38754..39035
/rpt_family="Alu"
39049..39326
/rpt_family="Alu"
complement(39424)
/rpt_family="Alu"
40705..41194
/rpt_family="Alu"
complement(41667)
/rpt_family="Alu"
42128..42424
/rpt_family="Alu"
42475..42957
/rpt_family="Alu"
43412..43710
/rpt_family="Alu"
complement(43754)

Query Match          76.8%; Score 19.2; DB 8; Length 159613;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTAGGCTTTTCAGT 24
      ||||| ||||| ||||| |||||
Db 1712 AAAAAATCTGTAGGTTTCATT 1689

RESULT 45
AC003420 Homo sapiens genomic DNA, chromosome 11q, clone:RP11-614K12,
LOCUS AP003420
DEFINITION Homo sapiens genomic DNA
ACCESSION AP003420
VERSION AP003420.2 GI:15320509
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.

REFERENCE
1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in Database (2001)
2 (bases 1 to 160945)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (16-MAR-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Aug 27, 2001 this sequence version replaced gi:13383333.
COMMENT Location/Qualifiers
FEATURES
1..160945
/mol_type="Homo sapiens"
/genome="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/maps="11q"
/clone="RP11-614K12"

ORIGIN

Query Match          76.8%; Score 19.2; DB 8; Length 160945;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTAGGCTTTTCAGT 25
      ||||| ||||| ||||| |||||
Db 133001 AAAAAATCCAGGAGGCTTCAGTG 133024

RESULT 46
AC163997 161009 bp DNA linear HTG 22-JUN-2005
LOCUS AC163997
DEFINITION Mus musculus chromosome 8 clone RP24-273M3 map 8, *** SEQUENCING IN
PROGRESS ***, 4 unordered pieces.
ACCESSION AC163997
VERSION AC163997.2 GI:68131736
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 161009)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 8, clone RP24-273M3
Unpublished
2 (bases 1 to 161009)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepl,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JUN-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 161009)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepl,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataranan,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS

```

JOURNAL Submitted (22-JUN-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jun 22, 2005 this sequence version replaced gi:67763997. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Broad Institute of MIT and Harvard  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@broad.mit.edu  
----- Project Information  
Center project name: L32316  
Center clone name: 273\_M\_3

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. \* as soon as it is available and the accession number will be preserved.

\* 1 21232: contig of 21232 bp in length  
\* 21233 21332: gap of unknown length  
\* 21333 91315: contig of 69983 bp in length  
\* 91316 91415: gap of unknown length  
\* 91416 145954: contig of 54539 bp in length  
\* 145955 146054: gap of unknown length  
\* 146055 161009: contig of 14955 bp in length.

FEATURES  
source  
1. .161009 Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="8"  
/map="8"  
/clone="RP24-273M3"  
/clone\_lib="RPCI-24 Male Mouse BAC"  
21233..21332  
/estimated\_length=unknown  
91316..91415  
/estimated\_length=unknown  
145955..146054  
/estimated\_length=unknown

ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 161009;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTCAGTG 25  
||||| ||| ||||| |||||  
Db 69604 AAAACAACTGAGGCTTTCAGTG 69627

RESULT 47  
AP000356/c 163795 bp DNA linear PRI 16-JUN-1999  
LOCUS Homo sapiens genomic DNA, chromosome 22q11.2, clone KB1995A5.  
DEFINITION AP000356  
ACCESSION AP000356.1 GI:5103019  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 163795)  
Shimizu,N.  
TITLE Human DNA sequence from clone KB1995A5 on chromosome 22q11.2  
JOURNAL Published Only in DataBase (1999)  
REFERENCE 2 (bases 1 to 163795)

AUTHORS Shimizu,N.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUN-1999) Nobuyoshi Shimizu, Keio university, school of medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan (E-mail: shimizu@mb.med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)  
COMMENT This is a complete sequence of the insert of KB1995A5 clone. The proximal adjacent clone is KB1896H10 (Acc.#AP000355) with 28999-bp overlapping. The distal adjacent clone is N95F10 (Acc.#AP000357) with 9742-bp overlapping.  
FEATURES  
Location/Qualifiers  
1. .163795  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/map="22q11.2"  
/clone="KB1995A5"  
601..1212  
/rpt\_family="ALU"  
2262..2556  
/rpt\_family="ALU"  
complement(2631..2808)  
/rpt\_family="L1"  
complement(3926..4211)  
/rpt\_family="ALU"  
complement(5143..5442)  
/rpt\_family="L1"  
8009..8291  
/rpt\_family="ALU"  
complement(9028..9324)  
/rpt\_family="ALU"  
10207..10469  
/rpt\_family="ALU"  
13212..13500  
/rpt\_family="ALU"  
13522..13782  
/rpt\_family="ALU"  
14972..15263  
/rpt\_family="ALU"  
complement(16821..17054)  
/rpt\_family="ALU"  
19386..19669  
/rpt\_family="ALU"  
20189..20429  
/rpt\_family="ALU"  
21005..21126  
/rpt\_family="ALU"  
21278..21587  
/rpt\_family="ALU"  
complement(21639..21859)  
/rpt\_family="ALU"  
22062..22294  
/rpt\_family="ALU"  
22353..22493  
/rpt\_family="L1"  
22667..22920  
/rpt\_family="ALU"  
complement(23138..23416)  
/rpt\_family="ALU"  
complement(24062..24373)  
/rpt\_family="ALU"  
24825..25114  
/rpt\_family="ALU"  
complement(25386..25553)  
/rpt\_family="ALU"  
complement(25849..25975)  
/rpt\_family="ALU"  
26752..27137  
/rpt\_family="ALU"  
complement(27233..27532)  
/rpt\_family="ALU"  
27970..28250  
/rpt\_family="ALU"

```
repeat_region /rpt_family="ALU"
29418. .29707
repeat_region /rpt_family="ALU"
complement(30743. .31449)
repeat_region /rpt_family="ALU"
complement(31594. .31881)
repeat_region /rpt_family="ALU"
32106. .32347
repeat_region /rpt_family="ALU"
33680. .33945
repeat_region /rpt_family="ALU"
37041. .37337
repeat_region /rpt_family="ALU"
complement(37911. .38207)
repeat_region /rpt_family="ALU"
complement(38286. .38560)
repeat_region /rpt_family="ALU"
complement(41817. .42107)
repeat_region /rpt_family="ALU"
44801. .45082
repeat_region /rpt_family="ALU"
complement(45132. .45223)
repeat_region /rpt_family="ALU"
45524. .45614
repeat_region /rpt_family="MER30"
46248. .46442
repeat_region /rpt_family="MIR"
complement(48671. .48806)
repeat_region /rpt_family="ALU"
49365. .49531
repeat_region /rpt_family="MER3"
complement(52097. .52183)
repeat_region /rpt_family="MLT1"
complement(52391. .52711)
repeat_region /rpt_family="ALU"
complement(52745. .53004)
repeat_region /rpt_family="ALU"
complement(53028. .53327)
repeat_region /rpt_family="ALU"
complement(53372. .53689)
repeat_region /rpt_family="ALU"
54030. .54282
repeat_region /rpt_family="ALU"
complement(54365. .54437)
repeat_region /rpt_family="ALU"
54541. .55027
repeat_region /rpt_family="ALU"
55398. .55449
repeat_region /rpt_family="ALU"
complement(56323. .56399)
repeat_region /rpt_family="MER42"
complement(57185. .57462)
repeat_region /rpt_family="ALU"
complement(57966. .58242)
repeat_region /rpt_family="ALU"
complement(58582. .58851)
repeat_region /rpt_family="ALU"
59776. .60056
repeat_region /rpt_family="ALU"
complement(60648. .60821)
repeat_region /rpt_family="ALU"
complement(60861. .61300)
repeat_region /rpt_family="ALU"
61779. .62664
repeat_region /rpt_family="MER21"
62354. .62627
repeat_region /rpt_family="MER39"
62703. .62991
repeat_region /rpt_family="ALU"
complement(63120. .63407)
repeat_region /rpt_family="ALU"
63780. .64061
repeat_region /rpt_family="ALU"
```

```
repeat_region 64794. .65072
/rpt_family="ALU"
repeat_region 65255. .65482
/rpt_family="ALU"
repeat_region 65910. .66157
/rpt_family="ALU"
repeat_region complement(66844. .67133)
/rpt_family="ALU"
repeat_region 67645. .67943
/rpt_family="ALU"
repeat_region complement(68043. .68339)
/rpt_family="ALU"
repeat_region complement(68550. .68851)
/rpt_family="ALU"
repeat_region 69601. .69888
/rpt_family="ALU"
repeat_region 69979. .70259
/rpt_family="ALU"
repeat_region 71892. .72253
/rpt_family="L1"
repeat_region 72368. .72636
/rpt_family="ALU"
repeat_region 73230. .73512
/rpt_family="ALU"
repeat_region 73847. .74133
/rpt_family="ALU"
repeat_region complement(75026. .75305)
/rpt_family="ALU"
repeat_region complement(78843. .79128)
/rpt_family="ALU"
repeat_region complement(82429. .82707)
/rpt_family="ALU"
repeat_region complement(83392. .83666)
/rpt_family="ALU"
repeat_region 85767. .86104
/rpt_family="ALU"
repeat_region complement(88527. .88684)
/rpt_family="MLT1"
repeat_region complement(88766. .88888)
```

Query Match 76.8%; Score 19.2; DB 8; Length 163795;

Best Local Similarity 87.5%; Pred. NO. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTAGGCTTTCAGTG 25

Db 70436 AAAAATCTCTAGCTTGTATTG 70413

RESULT 48

AC009048/c

LOCUS

DEFINITION

AC009048

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

AC009048 163844 bp DNA linear PRI 28-JUN-2001  
Homo sapiens chromosome 16 clone RP11-211H13, complete sequence.

AC009048.6 GI:14573673

HTG. Homo sapiens (human)

AC009048.6 Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens

HTG. Homo sapiens



JOURNAL Submitted (28-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 28, 2001 this sequence version replaced gi:7689878. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu

Quality: Phrap Quality >40 99.7% of Sequence; Estimated Total Number of Errors is 0.4.

STS Content: SHGC-52683 G36700 SHGC-60499 G36838.

FEATURES  
source Location/Qualifiers  
1..163844  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-211H13"

ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 163844;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
|||||  
Db 59236 AAAAAATCTGTAGGCTTTTCATT 59213  
|||||

RESULT 49  
AC019081/c 163865 bp DNA linear HTG 30-DEC-1999  
LOCUS Homo sapiens clone RP11-340P24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 11  
DEFINITION unordered pieces.  
AC019081  
VERSION AC019081.1 GI:6648469  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 163865)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 163865)  
Waterston,R.H.  
Direct Submission  
Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
Center project name: H NH0340P24.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 2004: contig of 2004 bp in length  
2005 2022: gap of unknown length  
2023 5849: contig of 3827 bp in length  
5867: gap of unknown length  
5868 9538: contig of 3671 bp in length  
9539 16258: gap of unknown length  
16259 16276: gap of unknown length  
16277 22543: contig of 6267 bp in length  
22544 22561: gap of unknown length  
22562 29557: contig of 6996 bp in length

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Submitted (28-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jun 28, 2001 this sequence version replaced gi:7689878. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu

Quality: Phrap Quality >40 99.7% of Sequence; Estimated Total Number of Errors is 0.4.

STS Content: SHGC-52683 G36700 SHGC-60499 G36838.

FEATURES  
source Location/Qualifiers  
1..163844  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-211H13"

ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 163844;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
|||||  
Db 59236 AAAAAATCTGTAGGCTTTTCATT 59213  
|||||

RESULT 50  
AC147169/c 165948 bp DNA linear VRT 23-OCT-2004  
LOCUS Xenopus tropicalis clone CH216-149115, complete sequence.  
DEFINITION AC147169  
ACCESSION AC147169  
VERSION AC147169.2 GI:54606892  
KEYWORDS HTG.  
SOURCE Xenopus tropicalis (Silurana tropicalis)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 165948)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Unpublished  
2 (bases 1 to 165948)  
DOE Joint Genome Institute.  
Direct Submission  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
CONSRTM  
TITLE

Submitted (05-NOV-2003) Production Genomics Facility, DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA  
3 (bases 1 to 165948)  
Stanford Human Genome Center.  
DOE Joint Genome Institute  
Direct Submission

JOURNAL Submitted (23-OCT-2004) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Oct 23, 2004 this sequence version replaced gi:38175822. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center www-shgc.stanford.edu

Quality: Phrap Quality >=40 100% of Sequence; Estimated Total Number Of Errors is 0.

FEATURES Location/Qualifiers

source 1..165948

/organism="Xenopus tropicalis"

/mol\_type="genomic DNA"

/db\_xref="taxon:8364"

/clone="CH216-149115"

ORIGIN

Query Match 76.8%; Score 19.2; DB 5; Length 165948;

Best Local Similarity 87.5%; Pred No. 1.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

Db 36239 AAAAACTCCTGTAGGCTGCCAGT 36216

Search completed: February 3, 2006, 21:26:53

Job time : 875 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-20

Perfect score: 25

Sequence: 1 aaaaactcctgtaggcttcagtg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21.\*

1: Geneseqn1980s.\*

2: Geneseqn1980s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	74868	13	ADV34991 Murine CD
C 2	19.8	79.2	599	13	Adq55481 Novel can
C 3	19.8	79.2	2243	13	Adq99098 GI to S p
C 4	19.2	76.8	439	10	Abt22339 Breast ca
C 5	19.2	76.8	545	13	Adul4171 Solid tum
C 6	19.2	76.8	562	4	Aak12378 Human bra
C 7	19.2	76.8	562	4	Abt37712 Human liv
C 8	19.2	76.8	562	6	Abt37712 Human liv
C 9	19.2	76.8	2527	11	Adm02046 Human gen
C 10	19.2	76.8	2527	10	Adc35122 Human bre
C 11	19.2	76.8	2587	12	Adn03792 Antipsori
C 12	19.2	76.8	2587	12	Adp13389 Renal cel
C 13	19.2	76.8	2587	13	Adp54303 Human PRO
C 14	19.2	76.8	2587	13	Adp24764 PRO polyP
C 15	19.2	76.8	2587	13	Adu06152 Novel bro
C 16	19.2	76.8	2587	14	Ady20501 DNA encod
C 17	19.2	76.8	2587	14	Ady14680 DNA encod
C 18	19.2	76.8	3886	10	Ades3710 Human pro
C 19	19.2	76.8	5569	5	Adl61917 Human ova

Abv24840 Human pro	5	5570	76.8	19.2	5	ABV24840
Aak84729 Human imm	4	48037	76.8	19.2	4	AAK84729
Aak85983 Human imm	4	48037	76.8	19.2	4	AAK85983
Aak84730 Human imm	4	48045	76.8	19.2	4	AAK84730
Aak85984 Human imm	4	48045	76.8	19.2	4	AAK85984
Ad433112 Human mit	10	466	74.4	18.6	10	ADD33112
Ad48290 Human enz	8	39982	74.4	18.6	8	AD48290
Adl13771 Osteoearth	10	201766	74.4	18.6	10	ADL13771
Aac36832 Arabidops	3	1175	73.6	18.4	3	AAC36832
Aac42714 Arabidops	3	2286	73.6	18.4	3	AAC42714
Adn74670 Thale cre	12	2286	73.6	18.4	12	ADN74670
Abn22936 Human ORF	6	463	72.8	18.2	6	ABN22936
Continuation (2 of	10	110000	72.8	18.2	10	ADH10017_1
Adg97266 Human can	12	110000	72.8	18.2	12	ADQ97266_0
Continuation (2 of	12	110000	72.8	18.2	12	ADQ97266_1
Continuation (3 of	12	110000	72.8	18.2	12	ADQ97328_2
Abd33491 Human can	13	112241	72.8	18.2	13	ABD33491
Adl13676 Osteoearth	10	192839	72.8	18.2	10	ADL13676
Aan82201 Beta-amyl	1	1794	71.2	17.8	1	AAN82201
Continuation (4 of	6	ABQ69245_03	71.2	17.8	6	ABQ69245_03
Continuation (3 of	6	ABQ67197_02	71.2	17.8	6	ABQ67197_02
Aal47655 Specific	43	43	70.4	17.6	43	AAL47655
Aal47656 Specific	6	43	70.4	17.6	6	AAL47656
Adb50438 Primary r	10	457	70.4	17.6	10	ADB50438
Adc56904 Bacterial	13	925	70.4	17.6	13	ADC56904
Adx09691 Plant ful	13	1303	70.4	17.6	13	ADX09691
Ab111811 Drosophil	4	1368	70.4	17.6	4	AB111811
Aag99349 Frog bomb	2	1563	70.4	17.6	2	AAQ99349
Aas12704 Frog bomb	5	1563	70.4	17.6	5	AAS12704
Ades3485 Human gen	10	2962	70.4	17.6	10	ADE3485
Ab111810 Drosophil	4	5010	70.4	17.6	4	AB111810
Ab118535 Drosophil	4	5904	70.4	17.6	4	AB118535
Ab118534 Drosophil	4	8685	70.4	17.6	4	AB118534
Adz13273 Human can	14	99544	70.4	17.6	14	ADZ13273
Acn45034 Human gen	11	99588	70.4	17.6	11	ACN45034
Ado34435 Human SLI	12	110000	70.4	17.6	12	ADO34435_0
Continuation (2 of	12	110000	70.4	17.6	12	ADO34435_1
Abk83497 Human CDN	6	227968	70.4	17.6	6	ABK83497
Adq18538 Human sof	12	227968	70.4	17.6	12	ADQ18538
Adg97146 Human can	12	348101	70.4	17.6	12	ADG97146
Aak54030 Murine tr	4	356	68.8	17.2	4	AAK54030
Achl6777 Human adu	9	426	68.8	17.2	9	ACH16777
Adg79472 Novel can	13	538	68.8	17.2	13	ADG79472
Abn60093 Human can	6	576	68.8	17.2	6	ABN60093
Adc163078 Human col	14	627	68.8	17.2	14	ADC163078
Aah03914 Human CDN	4	781	68.8	17.2	4	AAH03914
Aahl3859 Human CDN	4	1233	68.8	17.2	4	AAH13859
Acas57585 Human adi	8	1293	68.8	17.2	8	ACA57585
Aah33825 Human col	4	1315	68.8	17.2	4	AAH33825
Adq20266 Human sof	12	1348	68.8	17.2	12	ADQ20266
Adc35833 Novel mou	12	1425	68.8	17.2	12	ADC35833
Aah98243 Human EST	4	1496	68.8	17.2	4	AAH98243
Acad29572 Prokaryot	8	1857	68.8	17.2	8	ACA29572
Adf76674 Novel hum	10	1860	68.8	17.2	10	ADF76674
Adq86627 Human tum	13	1884	68.8	17.2	13	ADQ86627
Adq85502 Human tum	13	1884	68.8	17.2	13	ADQ85502
Adg24287 Human sof	12	1956	68.8	17.2	12	ADG24287
Aai59067 Human pol	5	1971	68.8	17.2	5	AAI59067
Adq99290 DNA encod	5	1971	68.8	17.2	5	ADQ99290
Adb49050 Novel hum	9	1971	68.8	17.2	9	ADB49050
Adc138806 Rice stre	11	2000	68.8	17.2	11	ADC138806
Aas79442 DNA encod	5	2006	68.8	17.2	5	AAS79442
Aahl7147 Human CDN	4	2008	68.8	17.2	4	AAH17147
Abal06762 Human CDN	4	2022	68.8	17.2	4	ABA06762
Aas31650 cDNA enco	4	2022	68.8	17.2	4	AAS31650
Abv84099 Human pol	6	2022	68.8	17.2	6	ABV84099
Adq64806 Novel hum	12	2386	68.8	17.2	12	ADQ64806
Aahl6475 Human CDN	4	2491	68.8	17.2	4	AAH16475
Aak94708 Human ful	4	2491	68.8	17.2	4	AAK94708
Adl31708 Full leng	12	2491	68.8	17.2	12	ADL31708
Aai60853 Human pol	4	2794	68.8	17.2	4	AAI60853
Aac77585 Human ORF	3	2795	68.8	17.2	3	AAC77585
Aas60774 Human can	4	3376	68.8	17.2	4	AAS60774

```
c 93 17.2 68.8 3914 10 ADB62696
c 94 17.2 68.8 4263 3 AA113379
c 95 17.2 68.8 4263 3 AA114341
c 96 17.2 68.8 6138 6 ABV73472
c 97 17.2 68.8 6141 6 ABV73471
c 98 17.2 68.8 6586 5 AAS45005
c 99 17.2 68.8 7646 6 ABV73469
c 100 17.2 68.8 16570 6 ABV73470
c 101 17.2 68.8 108359 9 ADA13316
c 102 17.2 68.8 169998 6 AAD36511
c 103 17.2 68.8 197496 6 AEN85584
c 104 17.2 68.8 200620 12 ADO56277
c 105 17.2 68.8 209083 13 ABD32854
c 106 17.2 68.8 213300 14 ADX80726
c 107 17.2 68.8 195 4 AAK83806
c 108 17.2 68.8 199 4 AAK62949
c 109 17.2 68.8 288 5 AAH52267
c 110 17.2 68.8 412 11 ACL38127
c 111 17.2 68.8 419 5 AAS84713
c 112 17.2 68.8 425 5 AAS84710
c 113 17.2 68.8 430 9 ACH18048
c 114 17.2 68.8 519 5 ABV14658
c 115 17.2 68.8 549 8 ABZ19092
c 116 17.2 68.8 900 2 AAT67966
c 117 17.2 68.8 1090 13 ADR65002
c 118 17.2 68.8 1095 10 ADH82740
c 119 17.2 68.8 1421 10 ADC32313
c 120 17.2 68.8 1482 8 ACA48417
c 121 17.2 68.8 1690 8 ABT42930
c 122 17.2 68.8 1939 6 ABZ11684
c 123 17.2 68.8 1939 12 ADM44202
c 124 17.2 68.8 1955 11 ADM03271
c 125 17.2 68.8 2000 12 ADJ41452
c 126 17.2 68.8 2064 12 ADF88612
c 127 17.2 68.8 2186 12 ADF88614
c 128 17.2 68.8 2393 10 ADA52804
c 129 17.2 68.8 2409 11 ADM01794
c 130 17.2 68.8 2526 14 AEB94776
c 131 17.2 68.8 2766 5 AAS73820
c 132 17.2 68.8 2766 5 AAS91674
c 133 17.2 68.8 2776 10 ADE25649
c 134 17.2 68.8 2905 6 ABK84106
c 135 17.2 68.8 2905 8 ABZ23871
c 136 17.2 68.8 2905 10 ADI18901
c 137 17.2 68.8 2905 10 ADD18899
c 138 17.2 68.8 2905 10 ADF76968
c 139 17.2 68.8 2905 12 ADJ75032
c 140 17.2 68.8 2905 12 ADN04093
c 141 17.2 68.8 2905 13 ADN25811
c 142 17.2 68.8 2905 13 ADR14236
c 143 17.2 68.8 2905 13 ADS64453
c 144 17.2 68.8 2905 14 ADY15451
c 145 17.2 68.8 2905 14 AEA23545
c 146 17.2 68.8 2905 14 AEA36131
c 147 17.2 68.8 2911 12 ADI32351
c 148 17.2 68.8 2911 12 ACN39568
c 149 17.2 68.8 2921 8 ABZ35982
c 150 17.2 68.8 3028 14 AEA61091
```

## ALIGNMENTS

```
RESULT 1
ADV34991/c
ID ADV34991 standard; cDNA; 74868 BP.
XX
AC ADV34991;
XX
DT 10-FEB-2005 (first entry)
XX
DE Murine cDNA differentially expressed in the presence of valproate Seq67.
```

```
Adb62696 Human cDN
Aaa13379 Mouse WT5
Aaa14341 cDNA enco
Abv73472 Human cyt
Aas45005 cDNA enco
Abv73469 Human cyt
Abv73470 Human cyt
Ada13316 Human ffr
Aad36511 Human Her
Abn85584 Human EGF
Ado56277 Human pre
Abd32854 Human can
Adx80726 Human RAL
Aak83806 Human imm
Aak62949 Human imm
Aah52267 Human AFP
Acl38127 Rice stre
Aas84713 DNA enco
Aas84710 DNA enco
Achn18048 Human adu
Abv14658 Human pro
Abz19092 Group III
Aat67966 H. pylori
Adr65002 Cotton cD
Adh82740 Enterococ
Adc32313 Human nov
Aca48417 Prokaryot
Abt42930 Human neu
Abz11684 Human pol
Adm44202 Novel hum
Adm03271 Human cDN
Adj41452 Plant cDN
Adf88612 DNA of hu
Adf88614 DNA of hu
Ada52804 Human cod
Adm01794 Human cDN
Aeb94776 DNA enco
Aas73820 DNA enco
Aas91674 DNA enco
Ade25649 Human cDN
Abk84106 Human cDN
Abz23871 Human 2'-
Adi18901 Human dis
Adi18899 Human dis
Adf76968 Novel hum
Adj75032 Marker ge
Adn04093 Antipsori
Adr25811 Breast ca
Adr14236 Human NF-
Ags64453 Human 2'-
Ady15451 DNA enco
Aea23545 Human PRO
Aea36131 Human nuc
Adi32351 Human OAS
Acn39568 Tumour-as
Abz35982 Human sec
Aea61091 Human cLD
```

```
KW murine; mouse; valproate; ss; multi-parameter high throughput screening;
KW MPRTs; disease signature; neuropsychiatric; neurodegenerative;
KW schizophrenia; bipolar affective disorder; BAD; autism; Parkinson's;
KW Alzheimer's disease; neuroleptic; nootropic; antimanic; antidepressa.
OS Mus musculus.
PN US2003096264-A1.
XX 22-MAY-2003.
PD 18-JUN-2002; 2002US-00175523.
XX 18-JUN-2001; 2001US-0299151P.
PR 07-SEP-2001; 2001US-0317828P.
PR 25-SEP-2001; 2001US-035150P.
PR 14-NOV-2001; 2001US-0333047P.
PR 18-JAN-2002; 2002US-0349936P.
PR 04-MAR-2002; 2002US-0361834P.
XX (PSYC-) PSYCHIATRIC GENOMICS INC.
PA Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Laeng P;
PI Palfreyman M, Rajan P;
XX WPI; 2004-118903/12.
XX Identifying a compound that can treat disease or disorders, such as, a
XX neuropsychiatric disorder e.g., schizophrenia, or autism, comprises
XX determining the expression of one or more efficacy genes in a cell
XX contacted with the test compound.
XX Claim 9; SEQ ID NO 67; 39pp; English.
XX This invention relates to a novel screening method identified as a multi-
XX parameter high throughput screening (MPRTs) assay. Specifically, it
XX refers to an assay that utilizes the disease signature of a plurality of
XX specific genes associated with a particular disease, and identifies
XX differential expression between those cells taken from individuals
XX affected by that disease and those that are not affected. The present
XX invention then describes the screening of candidate pharmaceutical
XX compounds to identify those that have a potential therapeutic benefit for
XX the treatment of neuropsychiatric and neurodegenerative disorders
XX including schizophrenia, bipolar affective disorder (BAD) and autism, as
XX well as Parkinson's and Alzheimer's disease. Accordingly, the compounds
XX of this invention exhibit various activities including neuroleptic,
XX nootropic, antimanic and antidepressant. Furthermore, the screening
XX method used in MPRTs will be automated, such that a large number of test
XX compounds may be rapidly screened with a minimal amount of labour and
XX effort. This polynucleotide is the cDNA sequence of a gene that is
XX differentially expressed in mice in the presence of the therapeutic
XX compound valproate, given in an exemplification of the invention.
XX Sequence 74868 BP; 20285 A; 17222 C; 16442 G; 20919 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 25; DB 13; Length 74868;
Best Local Similarity 100.0%; Pred. No. 0.29; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
QY 1 AAAAAAATCCTGCTAGGCTTTCAGTG 25
Db 33482 AAAAAAATCCTGCTAGGCTTTCAGTG 33458
RESULT 2
ADQ55481
ID ADQ55481 standard; DNA; 599 BP.
XX
AC ADQ55481;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID6783.
```

```

XX canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX WO2004063324-A2.
XX 29-JUL-2004.
XX
XX 05-MAY-2003; 2003WO-US013853.
XX
XX 03-MAY-2002; 2002US-0377240P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PFIZ ) PFIZER PROD INC.
XX
XX Diggins JC, Porter M, Wei T;
XX WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
XX Claim 1; SEQ ID NO 6783; 4lpp; English.
XX
XX This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northern blots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX
XX Sequence 599 BP; 182 A; 96 C; 99 G; 203 T; 0 U; 19 Other;
SQ
Query Match 79.2%; Score 19.8; DB 13; Length 599;
Best Local Similarity 91.3%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 290 AAAAACTCCTGTAGGCTATCAGT 312

RESULT 3
ADR99098/c
ID ADR99098 standard; DNA; 2243 BP.

```

```

XX ADR99098;
XX AC
XX 02-DEC-2004 (first entry)
XX DT
XX DE
XX G1 to S phase transition 1, GSPT1, coding sequence, SEQ ID 104.
XX KW
XX Cytostatic; breast cancer; cancer; human; gene; ds; GSPT1.
XX OS
XX Homo sapiens.
XX WO2004078035-A2.
XX PN
XX 16-SEP-2004.
XX PD
XX 27-FEB-2004; 2004WO-US007268.
XX PF
XX 28-FEB-2003; 2003US-0450655P.
XX PR
XX (FARB ) BAYER PHARM CORP.
XX PA
XX Eveleigh D, Bigwood D;
XX PI
XX WPI; 2004-653556/63.
XX DR P-PSDB; ADR99225.
XX DR GENBANK; BE906054.
XX
XX Diagnosing breast cancer comprises comparing the level of expression of
PT genes or gene products in a first biological sample taken from a patient
PT with that in a normal patient sample.
XX
XX Claim 2; SEQ ID NO 104; 53pp; English.
XX
XX The present invention relates to a method (M1) for diagnosing breast
CC cancer in a patient. The method comprises comparing the level of
CC expression of one or more genes or gene products in a biological sample
CC from the patient with that in a normal patient sample, where a difference
CC in the gene expression in the first sample compared to that in the second
CC sample is a diagnostic of the disease. Also claimed are: method (M2) for
CC distinguishing between normal and disease tissues; method (M3) for
CC monitoring the response of a breast cancer patient to treatment with an
CC anti-cancer agent; method (M4) for identifying a compound for treating
CC breast cancer; and an array for distinguishing between normal and disease
CC tissues comprising two or more probes corresponding to genes selected
CC from ADR98995-ADR99121 or comprising two or more polypeptides selected
CC from ADR99122-ADR99248. In M1 and M2 the genes are selected from ADR98995
CC -ADR99121 and the gene products are polypeptides selected from ADR99122-
CC ADR99248. M1 is useful for diagnosing breast cancer. M2 and the array are
CC useful for distinguishing between normal and disease tissue. M3 is useful
CC for monitoring the response of a breast cancer patient to treatment with
CC an anti-cancer agent. M4 is useful for identifying a compound for
CC treating breast cancer. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 2243 BP; 661 A; 443 C; 525 G; 549 T; 0 U; 65 Other;
SQ
Query Match 79.2%; Score 19.8; DB 13; Length 2243;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 2095 AAAAAATTCCTAGGTTTTCANT 2072

RESULT 4
ABT22339/c
ID ABT22339 standard; DNA; 439 BP.
XX AC
XX ABT22339;
XX DT 16-APR-2003 (first entry)

```

XX Breast cancer marker gene SEQ ID No 712.  
DE  
XX  
XX Cytostatic; vaccine; breast cancer marker gene; breast mass; immunogen;  
KW chemotherapy; tumour burden; bait protein; two-hybrid; three-hybrid;  
KW surrogate marker gene; pharmacodynamic marker gene; transgenic animal;  
KW human; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200285298-A2.  
PN  
XX  
XX 31-OCT-2002.  
PD  
XX  
XX 19-APR-2002; 2002WO-US012612.  
PF  
XX  
XX 20-APR-2001; 2001US-0285163P.  
PR  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX  
XX Lillie J, Palermo A, Wang Y, Steinmann K, Elias J, Mertens M;  
PI  
XX  
XX WPI; 2003-093053/08.  
DR  
XX  
XX Novel isolated polypeptide encoded by breast cancer marker gene, useful  
PT for diagnosing, staging, monitoring, prognosing and treating diseases  
PT associated with breast cancer.  
PT  
XX  
XX Disclosure; Page 189; 725pp; English.  
PS  
XX  
XX The invention relates to an isolated polypeptide encoded by a breast  
CC cancer marker gene comprising any of 141/21-805 nucleotide sequences,  
CC given in the specification. The methods of the invention are useful for  
CC diagnosing patients having an identified breast mass or symptoms  
CC associated with breast cancer, to diagnose breast cancer or its  
CC precursors, and for monitoring the efficacy of treatment of a breast  
CC cancer patient (e.g. efficacy of chemotherapy). The methods are also  
CC useful for evaluating a patient before, after or during therapy, to  
CC evaluate the reduction in a tumour burden. The breast cancer marker gene  
CC proteins are useful as immunogens for raising antibodies, by immunising a  
CC mammal with a breast cancer marker protein. The marker proteins are  
CC useful as bait proteins in a two-hybrid or three-hybrid assay, to  
CC identify other proteins which bind to or interact with the marker  
CC proteins. The breast cancer marker genes are useful as surrogate marker  
CC genes for one or more disorders, disease states or conditions leading to  
CC disease states, in particular, breast cancers. The breast cancer marker  
CC genes are useful as pharmacodynamic marker genes. An antibody which  
CC selectively binds to a protein of a breast cancer marker gene is useful  
CC for treating cancers, particularly breast cancers. The host cell of the  
CC invention is useful for producing non-human transgenic animals. This  
CC polynucleotide sequence represents one of the breast cancer marker genes  
CC of the invention  
XX  
XX Sequence 439 BP; 111 A; 92 C; 98 G; 125 T; 0 U; 13 Other;  
SQ  
Query Match 76.8%; Score 19.2; DB 10; Length 439;  
Best Local Similarity 87.5%; Pred. No. 71;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24  
Db 166 AAAAAATTCCTGTAGGTTTCATT 143  
RESULT 5  
ADU14171/C  
ID ADU14171 standard; DNA; 545 BP.  
XX  
XX ADU14171;  
AC  
XX  
XX 27-JAN-2005 (first entry)  
DT  
XX  
XX Solid tumour prognosis gene seqid 4610.  
DE

XX  
KW  
XX  
XX Cytostatic; gene therapy; expression profile; solid tumour;  
KW peripheral blood mononuclear cell; PBMC; prognosis; ds.  
XX  
OS Unidentified.  
XX  
XX WO2004097052-A2.  
PN  
XX  
XX 11-NOV-2004.  
PD  
XX  
XX 29-APR-2004; 2004WO-US013587.  
PF  
XX  
XX 29-APR-2003; 2003US-0466067P.  
PR  
XX  
XX 23-JAN-2004; 2004US-0538246P.  
PR  
XX  
XX (AMHP ) WYETH.  
PA  
XX  
XX (STRA/) STRAHS A.  
PA  
XX  
XX Strahs A, Trepicchio WL, Burczynski ME, Twine NC, Slonim DK;  
PI  
XX  
XX Immerman F, Dörner AJ;  
PI  
XX  
XX WPI; 2004-804779/79.  
DR  
XX  
XX A method, useful for prognosing and treating solid tumor, comprises  
PT comparing an expression profile of a gene expressed in peripheral blood  
PT mononuclear cells to a reference expression profile of a gene.  
PT  
XX  
XX Disclosure; Page; 111pp; English.  
PS  
XX  
XX The invention describes a method comprising comparing an expression  
CC profile of at least one gene in a peripheral blood sample of a patient to  
CC at least one reference expression profile of the at least one gene, where  
CC the patient has a solid tumour, and each of the gene is differentially  
CC expressed in peripheral blood mononuclear cells (PBMCs) of a first class  
CC of patients as compared to PBMCs of a second class of patients, where  
CC both the first and second classes of patients have the solid tumour, and  
CC each of the first and second classes is a subcluster formed by an  
CC unsupervised clustering analysis of gene expression profiles in PBMCs of  
CC a population of patients who have the solid tumour, and where the  
CC majority of the first class of patients has a first clinical outcome, and  
CC the majority of the second class of patients has a second clinical  
CC outcome. Also described are: a system comprising (i) a memory or a  
CC storage medium including data that represent an expression profile of at  
CC least one gene in a peripheral blood sample of a patient who has a solid  
CC tumour, (ii) at least another storage medium including data that  
CC represent at least one reference expression profile of the gene, (iii) a  
CC program capable of comparing the expression profile to the reference  
CC expression profile, and (iv) a processor capable of executing the  
CC program, where expression levels of the gene in peripheral blood  
CC mononuclear cells of patients who have the solid tumour correlate with  
CC clinical outcomes of the patients; and a nucleic acid or protein array  
CC comprising concentrated probes for solid tumour prognosis genes, where  
CC each of the solid tumour prognosis genes is differentially expressed in  
CC PBMCs of a first class of patients as compared to PBMCs of a second class  
CC of patients, where both the first and second classes of patients have a  
CC solid tumour, and where the first class of patients has a first clinical  
CC outcome, and the second class of patients has a second clinical outcome.  
CC The method, system, and array are useful for prognosing and treating  
CC solid tumours. This sequence represents a solid tumour prognosis gene of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_ftc\_sequences.  
XX  
XX Sequence 545 BP; 145 A; 116 C; 97 G; 159 T; 0 U; 28 Other;  
SQ  
Query Match 76.8%; Score 19.2; DB 13; Length 545;  
Best Local Similarity 87.5%; Pred. No. 73;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24  
Db 452 AAAAAATTCCTGTAGGTTTCATT 429

RESULT 6  
AAK12378/c  
ID AAK12378 standard; DNA; 562 BP.  
XX  
XX AAK12378;  
AC  
XX  
XX 05-NOV-2001 (first entry)  
DT  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 12369.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
KW ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157275-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US000667.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 12369; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention  
XX  
XX Sequence 562 BP; 202 A; 108 C; 133 G; 119 T; 0 U; 0 Other;  
SQ  
Query Match 76.8%; Score 19.2; DB 4; Length 562;  
Best Local Similarity 87.5%; Pred. No. 74;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAAGCTCTGTAGGCTTTTCAGTG 25  
Db 29 AAAAAGCTCTGTATGCTTTGATTG 6  
RESULT 7  
ABS37712/c  
ID ABS37712 standard; DNA; 562 BP.  
XX  
XX ABS37712;  
AC  
XX  
XX 25-FEB-2003 (first entry)  
DT  
XX  
DE Human liver single exon probe, SEQ ID NO 12702.  
XX  
XX Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;

KW coronary heart disease; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157273-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US000664.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-488998/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX  
XX Claim 1; SEQ ID NO 12702; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/ fragments). The probe hybridizes at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABS25011-ABS51005 represent human  
XX liver single exon nucleic acid probes of the invention. Note: The  
XX sequence information for this patent does not appear in the printed  
XX specification but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 562 BP; 202 A; 108 C; 133 G; 119 T; 0 U; 0 Other;  
SQ  
Query Match 76.8%; Score 19.2; DB 4; Length 562;  
Best Local Similarity 87.5%; Pred. No. 74;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAAGCTCTGTAGGCTTTTCAGTG 25  
Db 29 AAAAAGCTCTGTATGCTTTGATTG 6  
RESULT 8  
ABS12106/c  
ID ABS12106 standard; DNA; 562 BP.  
XX  
XX ABS12106;  
AC  
XX  
XX 19-AUG-2002 (first entry)  
DT  
XX  
DE Human genome-derived single exon probe from lung SEQ ID NO 12097.  
XX  
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;



KW hyaline membrane disease.

XX Homo sapiens.

OS WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples.

XX Claim 1; SEQ ID NO 12097; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioloeyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 562 BP; 202 A; 108 C; 133 G; 119 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 562;

Best Local Similarity 87.5%; Pred. No. 74;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

DB 2405 AAAAAATTCCTGTAGGTTTCATT 2382

RESULT 10

ID ADC35122/c

XX ADC35122 standard; cDNA; 2587 BP.

XX AC ADC35122;

XX 18-DEC-2003 (first entry)

XX

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 25

DB 29 AAAAACTCCTGTATGCTTTGATTG 6

RESULT 9

ID ADM02046/c

XX ADM02046 standard; cDNA; 2527 BP.

XX AC ADM02046;

XX 20-MAY-2004 (first entry)

XX Human cDNA of the invention SEQ ID NO:731.

XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.

XX Homo sapiens.

XX EPI1347046-A1.

XX 24-SEP-2003.

XX 12-APR-2002; 2002EP-00008400.

XX 22-MAR-2002; 2002JP-00137785.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S; Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Izrie R, Tamechika I; Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-723558/69.

XX P-PSDB; ADM04489.

XX New polynucleotides and polypeptides are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy.

XX Claim 1; SEQ ID NO 731; 305pp; English.

XX The invention relates to a novel human polynucleotide and the encoded polypeptide. A polynucleotide of the invention may have a use in gene therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful as a primer for synthesizing the polynucleotide or as a probe for detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are useful in gene therapy, for developing a diagnostic marker or medicines for regulating their expression and activity, or as a target of gene therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides are useful as pharmaceutical agents. The present sequence represents a cDNA sequence of the invention.

XX Sequence 2527 BP; 705 A; 589 C; 642 G; 591 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 11; Length 2527;

Best Local Similarity 87.5%; Pred. No. 93;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

DB 2405 AAAAAATTCCTGTAGGTTTCATT 2382

RESULT 10

ID ADC35122/c

XX ADC35122 standard; cDNA; 2587 BP.

XX AC ADC35122;

XX 18-DEC-2003 (first entry)

XX

```
DE Human breast cancer antigen polynucleotide seq id 6.
KW breast cancer; breast cancer diagnosis; breast cancer antigen; gene; ss.
XX
OS Homo sapiens.
XX
PN US2003108888-A1.
XX
PD 12-JUN-2003.
XX
PF 15-MAY-2002; 2002US-00146473.
XX
PR 15-MAY-2001; 2001US-0291150P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Scanlan MJ, Gout I, Stockert E, Old LJ, Cure A, Chen Y;
XX
DR WPI; 2003-829397/77.
DR P-PSDB; ADC35080.
XX
PT Diagnosing breast cancer in subject by obtaining biological sample from
PT subject, contacting sample with breast cancer-associated polypeptides,
PT determining specific binding between polypeptides and agents in sample.
XX
PS Claim 1; SEQ ID NO 6; 173pp; English.
XX
XX The invention describes a method of diagnosing breast cancer in subject
XX comprising contacting biological sample from subject with at least two
XX different breast cancer-associated polypeptides (I) encoded by nucleic
XX acid molecules (II) comprising sequence chosen from 42 fully defined
XX sequences as given in specification, determining specific binding between
XX (I) and agents in sample, where presence of the binding is diagnostic for
XX breast cancer. The method is useful for diagnosing breast cancer in a
XX subject. The sample is blood, lymph node fluid or breast discharge fluid.
XX This sequence encodes a breast cancer antigen.
XX
SQ Sequence 2587 BP; 711 A; 608 C; 670 G; 598 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 10; Length 2587;
Best Local Similarity 87.5%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
DB 2463 AAAAAATTCCTGTAGGTTTTCATT 2440
RESULT 11
ADN03792/c
ID ADN03792 standard; cDNA; 2587 BP.
XX
AC ADN03792;
XX
DT 01-JUL-2004 (first entry)
XX
DE Antipsoriatic cDNA sequence #94.
XX
XX ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS Homo sapiens.
XX
FN WO2004028479-A2.
XX
PD 08-APR-2004.
XX
PF 25-SEP-2003; 2003WO-US030907.
XX
PR 25-SEP-2002; 2002US-0414006P.
XX
PA (GETH ) GENENTECH INC.
XX
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;
DE Human breast cancer antigen polynucleotide seq id 6.
KW breast cancer; breast cancer diagnosis; breast cancer antigen; gene; ss.
XX
OS Homo sapiens.
XX
PN US2003108888-A1.
XX
PD 12-JUN-2003.
XX
PF 15-MAY-2002; 2002US-00146473.
XX
PR 15-MAY-2001; 2001US-0291150P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Scanlan MJ, Gout I, Stockert E, Old LJ, Cure A, Chen Y;
XX
DR WPI; 2003-829397/77.
DR P-PSDB; ADC35080.
XX
PT Diagnosing breast cancer in subject by obtaining biological sample from
PT subject, contacting sample with breast cancer-associated polypeptides,
PT determining specific binding between polypeptides and agents in sample.
XX
PS Claim 1; SEQ ID NO 6; 173pp; English.
XX
XX The invention describes a method of diagnosing breast cancer in subject
XX comprising contacting biological sample from subject with at least two
XX different breast cancer-associated polypeptides (I) encoded by nucleic
XX acid molecules (II) comprising sequence chosen from 42 fully defined
XX sequences as given in specification, determining specific binding between
XX (I) and agents in sample, where presence of the binding is diagnostic for
XX breast cancer. The method is useful for diagnosing breast cancer in a
XX subject. The sample is blood, lymph node fluid or breast discharge fluid.
XX This sequence encodes a breast cancer antigen.
XX
SQ Sequence 2587 BP; 711 A; 608 C; 670 G; 598 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 10; Length 2587;
Best Local Similarity 87.5%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
DB 2463 AAAAAATTCCTGTAGGTTTTCATT 2440
RESULT 12
ADP13389/c
ID ADP13389 standard; DNA; 2587 BP.
XX
AC ADP13389;
XX
DT 26-AUG-2004 (first entry)
XX
DE Renal cell carcinoma differentially expressed gene #125.
XX
XX ds; diagnosis; non-blood disease; solid tumor; gene expression;
XX peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
XX head/neck cancer; differential expression.
XX
OS Homo sapiens.
XX
FN WO2004048933-A2.
XX
PD 10-JUN-2004.
XX
PF 21-NOV-2003; 2003WO-US037481.
XX
PR 21-NOV-2002; 2002US-0427982P.
PR 03-APR-2003; 2003US-0459782P.
XX
XX (AMHP ) WYETH.
XX
XX (TWIN/) TWINE N C.
XX
XX (BURC/) BURCZYNSKI M E.
XX
XX (TREP/) TREPICCHIO W L.
XX
XX (DORN/) DORNER A.
XX
XX (STOV/) STOVER J A.
XX
XX (SLON/) SLONI D K.
XX
XX Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
PI Sloni DK;
XX
XX WPI; 2004-460799/43.
XX
XX Diagnosing non-blood disease such as solid tumor, involves comparing
XX differential expression profile of specific genes in peripheral blood
XX sample of subject with reference expression profile of specific genes.
XX
XX Disclosure; SEQ ID NO 125; 350pp; English.
XX
XX The invention relate to a method of diagnosing (M1) non-blood disease
```



PT renal disease, or demyelinating diseases of the central or peripheral  
PT nervous system.

XX Claim 1; SEQ ID NO 1942; 2940pp; English.

XX The invention relates to a novel isolated nucleic acid and the PRO  
CC polypeptide encoded by it. A protein of the invention has  
CC antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,  
CC osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,  
CC antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide  
CC of the invention may have a use in gene therapy. The PRO polypeptide, its  
CC agonist, antagonist, or antibody that specifically binds to the  
CC polypeptide is useful for treating an immune related disorder such as  
CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an  
CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic  
CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune  
CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal  
CC disease, a demyelinating disease of the central or peripheral nervous  
CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,  
CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary  
CC disease, infectious or autoimmune chronic active hepatitis, primary  
CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,  
CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's  
CC disease, an autoimmune or immune-mediated skin disease, a bullous skin  
CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic  
CC disease, asthma, allergic rhinitis, atopic dermatitis, food  
CC hypersensitivity, urticaria, an immunologic disease of the lung,  
CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity  
CC pneumonitis, a transplantation associated disease, graft rejection or  
CC graft-versus-host disease. The present sequence encodes a PRO protein of  
CC the invention.

XX SQ Sequence 2587 BP; 711 A; 608 C; 670 G; 598 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 13; Length 2587;  
Best Local Similarity 87.5%; Pred. No. 93;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGT 24

||||| ||||||| |||||  
Db 2463 AAAAAAATCTCTAGGCTTTTCATT 2440

RESULT 15

ID ADU06152/c  
ADU06152 standard; DNA; 2587 BP.

AC ADU06152;

XX 27-JAN-2005 (first entry)

XX Novel bronchial cancer-associated human gene SeqID376.

DE bronchial' cancer; cytostatic; tumour-associated protein;  
KW cancer detection; metastasis; tumour; gene; ds; human.

XX Homo sapiens.

XX DE10316701-A1.

XX 04-NOV-2004.

XX 09-APR-2003; 2003DE-01016701.

XX 09-APR-2003; 2003DE-01016701.

XX (HINZ/) HINZMANN B.

PA (HERM/) HERMANN K.

PA (CAST/) HEIDEN CASTANOS-VELEZ E.

XX Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnewann H;  
PI Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;

XX

DR WPI; 2004-786403/78.

XX P-PSDB; ADU06639.

PT New nucleic acid, and derived proteins, useful for diagnosis of bronchial  
PT cancer and in screening for therapeutic and diagnostic agents.

XX Claim 1; SEQ ID NO 376; 1381pp; German.

XX This invention relates to a novel isolated nucleic acid associated with  
CC bronchial cancer comprising 489 defined sequences given in the  
CC specification. The invention may be useful for the production of  
CC compounds with a cytostatic activity through the inhibition of expression  
CC or activity of tumour-associated proteins. The novel DNA sequences and  
CC the proteins/peptides encoded by them are used for detecting bronchial  
CC cancer or determining the risk of developing it and to screen for  
CC specific binding partners of the DNA or protein sequences, where the  
CC binding partners are potentially useful as agents for treating or  
CC diagnosing bronchial cancer. The DNA or protein sequences can also be  
CC used for prognosis, detection of metastases and for secondary treatment  
CC (of tumours that have been stabilised or are no longer detectable).  
CC Detecting abnormal expression of the DNA sequences provides early  
CC diagnosis of bronchial cancers. The present sequence is that of a novel  
CC bronchial cancer-associated human gene sequence of the invention.

XX SQ Sequence 2587 BP; 711 A; 608 C; 670 G; 598 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 13; Length 2587;

Best Local Similarity 87.5%; Pred. No. 93;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGT 24

||||| ||||||| |||||  
Db 2463 AAAAAAATCTCTAGGCTTTTCATT 2440

RESULT 16

ADY20501/c

ID ADY20501 standard; DNA; 2587 BP.

XX AC ADY20501;

XX 05-MAY-2005 (first entry)

XX DNA encoding a PRO polypeptide, SEQ ID NO 6307.

XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;  
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;  
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;  
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;  
KW Antiallergic; ds; gene; diagnosis.

XX Homo sapiens.

XX WO2005016962-A2.

XX 24-FEB-2005.

XX 11-AUG-2004; 2004WO-US026249.

XX 11-AUG-2003; 2003US-0493546P.

XX (GETH ) GENENTECH INC.

XX Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;

XX WPI; 2005-182330/19.

XX New nucleic acid encoding PRO polypeptide, useful for diagnosing and  
PT treating an immune related disorder, e.g. systemic lupus erythematosus,  
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.  
XX Claim 1; SEQ ID NO 6307; 158pp; English.



KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX Homo sapiens.

XX WO200170979-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009126.

XX 21-MAR-2000; 2000US-0191031P.

XX 25-MAY-2000; 2000US-0207124P.

XX 15-JUN-2000; 2000US-0211940P.

XX 07-JUL-2000; 2000US-0216920P.

XX 25-JUL-2000; 2000US-0220661P.

XX 21-DEC-2000; 2000US-0257672P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lee J, Lillie J;

XX WPI; 2001-611502/70.

XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

XX Disclosure; SEQ ID NO 20129; 106pp; English.

XX The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, cancer polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 5569 BP; 1655 A; 1123 C; 1062 G; 1721 T; 0 U; 8 Other;

XX Query Match 76.8%; Score 19.2; DB 5; Length 5569;

XX Best Local Similarity 87.5%; Pred. No. 1.1e+02;

XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTGTAGGCTTTTCAGT 24

|||||

Db 317 AAAAAAATCTCTGTAGGCTTTTCATT 294

RESULT 20

ABV24840/c

ID ABV24840 standard; CDNA; 5570 BP.

XX AC ABV24840;

XX 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 24831.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX Pharmacogenomic marker; Gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX 16-MAR-2000; 2000US-0189862P.

XX 25-MAY-2000; 2000US-0207454P.

XX 09-JUN-2000; 2000US-0211314P.

XX 18-JUL-2000; 2000US-0219007P.

XX 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 4764-4766; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 5570 BP; 1655 A; 1124 C; 1062 G; 1721 T; 0 U; 8 Other;

XX Query Match 76.8%; Score 19.2; DB 5; Length 5570;

XX Best Local Similarity 87.5%; Pred. No. 1.1e+02;

XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTGTAGGCTTTTCAGT 24

|||||

Db 317 AAAAAAATCTCTGTAGGCTTTTCATT 294

RESULT 21

AAK84729/c

ID AAK84729 standard; DNA; 48037 BP.

XX AAK84729;

XX 07-NOV-2001 (first entry)

XX

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39541.  
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
cytostatic; gene therapy; vaccine; metastasis; ds.

DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39541.	PR	21-SEP-2000; 2000US-0234223P.
XX		PR	21-SEP-2000; 2000US-0234274P.
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	PR	25-SEP-2000; 2000US-0234997P.
XX	cytostatic; gene therapy; vaccine; metastasis; ds.	PR	25-SEP-2000; 2000US-0234998P.
OS	Homo sapiens.	PR	26-SEP-2000; 2000US-0235484P.
XX		PR	27-SEP-2000; 2000US-0235834P.
XX		PR	27-SEP-2000; 2000US-0235836P.
PN	WO200157182-A2.	PR	29-SEP-2000; 2000US-0236327P.
XX		PR	29-SEP-2000; 2000US-0236367P.
PD		PR	29-SEP-2000; 2000US-0236368P.
XX	09-AUG-2001.	PR	29-SEP-2000; 2000US-0236369P.
XX		PR	29-SEP-2000; 2000US-0236370P.
XX	17-JAN-2001; 2001WO-US001354.	PR	02-OCT-2000; 2000US-0236802P.
PR		PR	02-OCT-2000; 2000US-0237037P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237038P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237039P.
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000; 2000US-0237040P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239935P.
PR	16-MAR-2000; 2000US-0189874P.	PR	13-OCT-2000; 2000US-0239937P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0240960P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241221P.
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000; 2000US-0241785P.
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000; 2000US-0241786P.
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000; 2000US-0241808P.
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000; 2000US-0241809P.
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000; 2000US-0241826P.
PR	07-JUL-2000; 2000US-0216880P.	PR	20-OCT-2000; 2000US-0244617P.
PR	11-JUL-2000; 2000US-0217487P.	PR	01-NOV-2000; 2000US-0246474P.
PR	11-JUL-2000; 2000US-0217496P.	PR	08-NOV-2000; 2000US-0246475P.
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000; 2000US-0246476P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246477P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246478P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246523P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246524P.
PR	14-AUG-2000; 2000US-0225213P.	PR	08-NOV-2000; 2000US-0246525P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246526P.
PR	14-AUG-2000; 2000US-0225266P.	PR	08-NOV-2000; 2000US-0246527P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246528P.
PR	14-AUG-2000; 2000US-0225268P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0225270P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225447P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225757P.	PR	08-NOV-2000; 2000US-0246611P.
PR	14-AUG-2000; 2000US-0225758P.	PR	08-NOV-2000; 2000US-0246613P.
PR	14-AUG-2000; 2000US-0225759P.	PR	17-NOV-2000; 2000US-0249207P.
PR	18-AUG-2000; 2000US-0226279P.	PR	17-NOV-2000; 2000US-0249208P.
PR	22-AUG-2000; 2000US-0226681P.	PR	17-NOV-2000; 2000US-0249209P.
PR	22-AUG-2000; 2000US-0226868P.	PR	17-NOV-2000; 2000US-0249210P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249211P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249212P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249213P.
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000; 2000US-0249214P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249215P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249216P.
PR	01-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249217P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249218P.
PR	05-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249244P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249245P.
PR	06-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249264P.
PR	08-SEP-2000; 2000US-0231242P.	PR	17-NOV-2000; 2000US-0249265P.
PR	08-SEP-2000; 2000US-0231243P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0231244P.	PR	17-NOV-2000; 2000US-0249299P.
PR	08-SEP-2000; 2000US-0231413P.	PR	17-NOV-2000; 2000US-0249300P.
PR	08-SEP-2000; 2000US-0231414P.	PR	01-DEC-2000; 2000US-0250160P.
PR	08-SEP-2000; 2000US-0232080P.	PR	01-DEC-2000; 2000US-0250391P.
PR	08-SEP-2000; 2000US-0232081P.	PR	05-DEC-2000; 2000US-0251030P.
PR	12-SEP-2000; 2000US-0231968P.	PR	05-DEC-2000; 2000US-0251988P.
PR	14-SEP-2000; 2000US-0232397P.	PR	06-DEC-2000; 2000US-0251479P.
PR	14-SEP-2000; 2000US-0232398P.	PR	08-DEC-2000; 2000US-0251856P.
PR	14-SEP-2000; 2000US-0232399P.	PR	08-DEC-2000; 2000US-0251868P.
PR	14-SEP-2000; 2000US-0232400P.	PR	08-DEC-2000; 2000US-0251869P.
PR	14-SEP-2000; 2000US-0232401P.	PR	08-DEC-2000; 2000US-0251989P.
PR	14-SEP-2000; 2000US-0233063P.		
PR	14-SEP-2000; 2000US-0233064P.		
PR	14-SEP-2000; 2000US-0233065P.		



```
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 39541; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
XX SQ Sequence 48037 BP; 12892 A; 9814 C; 10466 G; 14865 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 4; Length 48037;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGAGGCTTTCACT 24
   ||||| ||||| ||||| |||||
Db 43155 AAAAAATCCTGAGGTTTTCATT 43132

RESULT 22
AAK85983/c
ID AAK85983 standard; DNA; 48037 BP.
AC AAK85983;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40795.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; db.
XX
XX Homo sapiens.
XX
XX W0200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
```

PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
PT  
XX  
XX Disclosure; SEQ ID NO 40795; 3071pp + Sequence Listing; English.  
PS  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to prevent,  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 48037 BP; 12892 A; 9814 C; 10466 G; 14865 T; 0 U; 0 Other;  
  
Query Match 76.8%; Score 19.2; DB 4; Length 48037;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAATCTCTAGGCTTCAGT 24  
Db 43155 AAAAATCTCTAGGCTTCATT 43132  
  
RESULT 23  
AAK84730/c  
ID AAK84730 standard; DNA; 48045 BP.  
XX  
AC AAK84730;  
XX  
DT 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39542.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-021680P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225759P.

14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 06-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 08-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231124P.  
 PR 08-SEP-2000; 2000US-0231124P.  
 PR 08-SEP-2000; 2000US-0231143P.  
 PR 08-SEP-2000; 2000US-0231143P.  
 PR 08-SEP-2000; 2000US-0231144P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.

08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX  
 DR  
 XX  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 PS Disclosure; SEQ ID NO 39542; 3071pp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 48045 BP; 12889 A; 9812 C; 10461 G; 14883 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 4; Length 48045;  
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 43163 AAAAAATTCCTGTAGGTTTTCATT 43140

RESULT 24	
AAK85984/c	
ID AAK85984 standard; DNA; 48045 BP.	
XX	
AC AAK85984;	
XX	
DT 07-NOV-2001 (first entry)	
XX	
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40796.	
XX	
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW cytostatic; gene therapy; vaccine; metastasis; ds.	
XX	
OS Homo sapiens.	
XX	
PN WO200157182-A2.	
XX	
PD 09-AUG-2001.	
XX	
PF 17-JAN-2001; 2001WO-US001354.	
XX	
31-JAN-2000; 2000US-0179065P.	PR 08-SEP-2000; 2000US-0232081P.
PR 04-FEB-2000; 2000US-0180628P.	PR 12-SEP-2000; 2000US-0231968P.
PR 24-FEB-2000; 2000US-0184664P.	PR 14-SEP-2000; 2000US-0232397P.
PR 02-MAR-2000; 2000US-0186350P.	PR 14-SEP-2000; 2000US-0232398P.
PR 16-MAR-2000; 2000US-0189874P.	PR 14-SEP-2000; 2000US-0232399P.
PR 17-MAR-2000; 2000US-0190076P.	PR 14-SEP-2000; 2000US-0232400P.
PR 18-APR-2000; 2000US-0198123P.	PR 14-SEP-2000; 2000US-0232401P.
PR 19-MAY-2000; 2000US-0205515P.	PR 14-SEP-2000; 2000US-0233063P.
PR 07-JUN-2000; 2000US-0209467P.	PR 14-SEP-2000; 2000US-0233064P.
PR 28-JUN-2000; 2000US-0214886P.	PR 14-SEP-2000; 2000US-0233065P.
PR 30-JUN-2000; 2000US-0215135P.	PR 21-SEP-2000; 2000US-0234223P.
PR 07-JUL-2000; 2000US-0216647P.	PR 21-SEP-2000; 2000US-0234274P.
PR 07-JUL-2000; 2000US-0216880P.	PR 25-SEP-2000; 2000US-0234997P.
PR 11-JUL-2000; 2000US-0217487P.	PR 25-SEP-2000; 2000US-0234998P.
PR 11-JUL-2000; 2000US-0217496P.	PR 26-SEP-2000; 2000US-0235484P.
PR 14-JUL-2000; 2000US-0218290P.	PR 27-SEP-2000; 2000US-0235834P.
PR 26-JUL-2000; 2000US-0220963P.	PR 27-SEP-2000; 2000US-0235836P.
PR 26-JUL-2000; 2000US-0220964P.	PR 29-SEP-2000; 2000US-0236327P.
PR 14-AUG-2000; 2000US-0224518P.	PR 29-SEP-2000; 2000US-0236367P.
PR 14-AUG-2000; 2000US-0224519P.	PR 29-SEP-2000; 2000US-0236368P.
PR 14-AUG-2000; 2000US-0225213P.	PR 29-SEP-2000; 2000US-0236369P.
PR 14-AUG-2000; 2000US-0225214P.	PR 29-SEP-2000; 2000US-0236370P.
PR 14-AUG-2000; 2000US-0225266P.	PR 02-OCT-2000; 2000US-0236802P.
PR 14-AUG-2000; 2000US-0225267P.	PR 02-OCT-2000; 2000US-0237037P.
PR 14-AUG-2000; 2000US-0225268P.	PR 02-OCT-2000; 2000US-0237038P.
PR 14-AUG-2000; 2000US-0225270P.	PR 02-OCT-2000; 2000US-0237039P.
PR 14-AUG-2000; 2000US-0225447P.	PR 02-OCT-2000; 2000US-0237040P.
PR 14-AUG-2000; 2000US-0225757P.	PR 13-OCT-2000; 2000US-0239935P.
PR 14-AUG-2000; 2000US-0225758P.	PR 13-OCT-2000; 2000US-0239937P.
PR 14-AUG-2000; 2000US-0225759P.	PR 20-OCT-2000; 2000US-0240960P.
PR 18-AUG-2000; 2000US-0226279P.	PR 20-OCT-2000; 2000US-0241221P.
PR 22-AUG-2000; 2000US-0226681P.	PR 20-OCT-2000; 2000US-0241785P.
PR 22-AUG-2000; 2000US-0226868P.	PR 20-OCT-2000; 2000US-0241786P.
PR 22-AUG-2000; 2000US-0227182P.	PR 20-OCT-2000; 2000US-0241787P.
PR 23-AUG-2000; 2000US-0227009P.	PR 20-OCT-2000; 2000US-0241808P.
PR 30-AUG-2000; 2000US-0228924P.	PR 20-OCT-2000; 2000US-0241809P.
PR 01-SEP-2000; 2000US-0229287P.	PR 01-NOV-2000; 2000US-0244617P.
PR 01-SEP-2000; 2000US-0229343P.	PR 08-NOV-2000; 2000US-0246474P.
PR 01-SEP-2000; 2000US-0229344P.	PR 08-NOV-2000; 2000US-0246475P.
PR 01-SEP-2000; 2000US-0229345P.	PR 08-NOV-2000; 2000US-0246476P.
PR 05-SEP-2000; 2000US-0229509P.	PR 08-NOV-2000; 2000US-0246477P.
PR 05-SEP-2000; 2000US-0229513P.	PR 08-NOV-2000; 2000US-0246478P.
PR 06-SEP-2000; 2000US-0230437P.	PR 08-NOV-2000; 2000US-0246523P.
PR 06-SEP-2000; 2000US-0230438P.	PR 08-NOV-2000; 2000US-0246524P.
PR 08-SEP-2000; 2000US-0231242P.	PR 08-NOV-2000; 2000US-0246525P.
PR 08-SEP-2000; 2000US-0231243P.	PR 08-NOV-2000; 2000US-0246526P.
PR 08-SEP-2000; 2000US-0231244P.	PR 08-NOV-2000; 2000US-0246527P.
PR 08-SEP-2000; 2000US-0231413P.	PR 08-NOV-2000; 2000US-0246528P.
PR 08-SEP-2000; 2000US-0231414P.	PR 08-NOV-2000; 2000US-0246532P.
PR 08-SEP-2000; 2000US-0232080P.	PR 08-NOV-2000; 2000US-0246609P.
	PR 08-NOV-2000; 2000US-0246610P.
	PR 08-NOV-2000; 2000US-0246611P.
	PR 08-NOV-2000; 2000US-0246613P.
	PR 17-NOV-2000; 2000US-0249207P.
	PR 17-NOV-2000; 2000US-0249208P.
	PR 17-NOV-2000; 2000US-0249209P.
	PR 17-NOV-2000; 2000US-0249210P.
	PR 17-NOV-2000; 2000US-0249211P.
	PR 17-NOV-2000; 2000US-0249212P.
	PR 17-NOV-2000; 2000US-0249213P.
	PR 17-NOV-2000; 2000US-0249214P.
	PR 17-NOV-2000; 2000US-0249215P.
	PR 17-NOV-2000; 2000US-0249216P.
	PR 17-NOV-2000; 2000US-0249217P.
	PR 17-NOV-2000; 2000US-0249218P.
	PR 17-NOV-2000; 2000US-0249244P.
	PR 17-NOV-2000; 2000US-0249245P.
	PR 17-NOV-2000; 2000US-0249264P.
	PR 17-NOV-2000; 2000US-0249265P.
	PR 17-NOV-2000; 2000US-0249297P.
	PR 17-NOV-2000; 2000US-0249299P.
	PR 17-NOV-2000; 2000US-0249300P.

```
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 40796; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK51921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 48045 BP; 12889 A; 9812 C; 10461 G; 14883 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 4; Length 48045;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
DB 43163 AAAAAATTCCTGTAGGTTTTCATT 43140
RESULT 25
ADD33112
ID ADD33112 standard; DNA; 466 BP.
XX AC ADD33112;
XX 15-JAN-2004 (first entry)
XX Human mitochondrial DNA sequence SEQ ID NO:883.
XX ds; human; array; mitochondrial; hybridisation; energy-metabolism;
KW mitochondrial disease; oxidative phosphorylation dysfunction;
KW oxidative stress; apoptosis; aging.
XX Homo sapiens.
XX
XX WO2003020220-A2.
XX
PD 13-MAR-2003.
```

```
XX 30-AUG-2002; 2002WO-US027886.
PF
XX 30-AUG-2001; 2001US-0316323P.
PR 31-AUG-2001; 2001CA-02356540.
XX (UYEM-) UNIV EMORY.
XX Wallace DC, Levy S, Kerstann K, Procaccio V;
PI WPI; 2003-300821/29.
XX
DR Array containing probes for genes involved in mitochondrial biology,
PT useful for determining mitochondrial biology gene expression profiles for
PT use in diagnosing pathologies and identifying biochemical pathways.
XX
PS Claim 2; SEQ ID NO 883; 201pp; English.
XX
CC The invention relates to a novel array comprising at least two isolated
CC nucleotide molecules, each molecule having a sequence capable of uniquely
CC hybridising to a nucleic acid molecule which is an expression product of
CC a gene involved in mitochondrial biology. The array comprises two or more
CC isolated nucleic acid molecules or spots, each molecule having a sequence
CC chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC of the invention is useful for determining an expression profile of a
CC mouse or human sample containing nucleic acid, by contacting the array
CC with the sample under conditions allowing selective hybridisation, and
CC measuring hybridisation of nucleic acid in the sample to the array to
CC produce an expression profile. The array is also useful for determining
CC an expression profile of a first labelled sample containing nucleic acid
CC relative to a second, differently labelled sample containing nucleic
CC acid. The second sample is a reference or a standard. An array is useful
CC for determining an expression profile diagnostic of an energy-metabolism-
CC related physiological condition. An array of the invention is useful for
CC determining mitochondrial biology gene expression profiles of organisms,
CC such as human, mice and closely related species, tissue and organs of
CC such organisms, which are useful for determining expression profiles
CC diagnostic of energy metabolism-related physiological conditions,
CC diagnosing such physiological conditions, identifying biochemical
CC pathways, genes, and mutations involved in such physiological conditions,
CC identifying therapeutic agents useful for preventing and/or treating such
CC physiological conditions, evaluating and/or monitoring the efficacy of
CC such therapies, and creating and identifying animal models of human
CC energy metabolism-related physiological conditions. An array is also
CC useful for defining expression signatures or profiles for mitochondrial
CC diseases, as well as distinguishing clinical disorders that result from
CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC apoptosis and aging. An array of the invention contains probes of genes
CC not previously recognised to participate in mitochondrial biology. The
CC sequences shown in ADD32231-ADD33223 represent human mitochondrial DNA
CC clones used to make the probes of the invention. Some sequences are not
CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC 1906, 2408 and 2643.
XX
SQ Sequence 466 BP; 86 A; 122 C; 124 G; 130 T; 0 U; 4 Other;
Query Match 74.4%; Score 18.6; DB 10; Length 466;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAACTCTGTAGGCTTTTCAGT 25
DB 328 AATATCATCTGTAGGCTTTTCGTG 352
RESULT 26
AAD48290
ID AAD48290 standard; DNA; 39982 BP.
XX AC AAD48290;
XX
XX 24-FEB-2003 (first entry)
XX
```

```
DE Human enzyme protein encoding gene.
XX
XX
XX Human; enzyme; drug screening; immune response; leukaemia; lymphoma;
KW pharmacogenomic; tissue typing; gene therapy; gene; ds.
XX
XX Homo sapiens.
XX
XX
XX Location/Qualifiers
FH 2201..36984
FT /tag= a
FT /product= "Human protein"
FT exon 2201..2367
FT /tag= b
FT intron 2368..3951
FT /tag= c
FT variation replace(2659,G)
FT /tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(3300,G)
FT /tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT exon 3952
FT /tag= f
FT intron 3953..6692
FT /tag= g
FT variation replace(4697, -)
FT /tag= h
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(4703, -)
FT /tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(4985,A)
FT /tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT exon 6693..6725
FT /tag= k
FT intron 6726..9189
FT /tag= l
FT variation replace(7837,T)
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(7874, -)
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(8299,C)
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(8916,A)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(8938,A)
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(8974,A)
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9182,T)
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT exon 9190..9254
FT /tag= t
FT intron 9255..9798
FT /tag= u
FT variation replace(9309,G)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9312,G)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9321,A)
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9539,C)
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9545,A)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9734,A)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT exon 9799..9907
FT /tag= ab
FT intron 9908..14148
FT /tag= ac
FT variation replace(9929,G)
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9996,C)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(10234,G)
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(10316,C)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(10565,G)
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(10622,C)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(10839,T)
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(11006,C)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(12853,T)
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13471,C)
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13500,A)
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13507,A)
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13554,C)
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13737,T)
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(13940,T)
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT exon 14149..14279
FT /tag= as
FT intron 14280..21593
FT /tag= at
FT variation replace(15427,A)
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(15486,C)
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(15648,G)
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(15650,T)
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT
```





PR 09-MAR-1999; 99US-0123548B.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142377P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.

```
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 73.6%; Score 18.4; DB 3; Length 1175;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AACTCCTGTAGGCTTTCAGT 24
    |||||
Db 367 AACTCCTGTAGGCTTTCAGT 348

RESULT 29
AAC42714/c
ID AAC42714 standard; DNA; 2286 BP.
XX
AC AAC42714;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36582.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
```

PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149829P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.

PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 22-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 73.6%; Score 18.4; DB 3; Length 2286;  
Best Local Similarity 95.0%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AACTCCTGTAGCGTTTCAGT 24  
|||||  
Db 1550 AACTCCTGTAGCGTTTCAGT 1531

RESULT 30  
ADN74670/c  
ID ADN74670 standard; cDNA; 2286 BP.  
XX AC ADN74670;  
XX 15-JUL-2004 (first entry)  
XX DE Thale cress cDNA repressed in E2Fa/Dpa expressing plants SeqID 2565.  
XX KW gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;  
KW growth regulator; animal feed product; thale cress;  
KW cell wall biosynthesis; nitrogen metabolism; carbon metabolism.  
XX OS Arabidopsis thaliana.  
XX PN WO2004035798-A2.  
XX PD 29-APR-2004.  
XX PF 20-OCT-2003; 2003WO-EP011658.  
XX PR 18-OCT-2002; 2002EP-00079408.  
XX PA (CROP-) CROPDESIGN NV.  
XX PI Inze D, De Veylder L, Vlieghe K;  
XX WPI; 2004-348466/32.  
XX P-PSDB; ADN74671.  
XX Altering plant characteristics, useful for producing plants for enzyme or  
XX pharmaceutical production comprising modifying in a plant, expression of  
XX one or more nucleic acids and/or modifying level or activity of one or  
XX more proteins.  
XX Claim 1; SEQ ID NO 2565; 134pp; English.

CC This invention relates to a novel method for altering one or more plant  
CC characteristics. Specifically, it refers to identifying genes that are up  
CC - or down-regulated in transgenic plants overexpressing the heterodimeric  
CC E2Fa/DPA transcription factor of Arabidopsis and using these sequences to  
CC alter plant characteristics accordingly. The present invention describes  
CC generating transgenic plants for the production of growth regulators,  
CC enzymes, therapeutics, pharmaceuticals and animal feed products, where  
CC the altered plant characteristics are selected from increased yield or  
CC biomass, enhanced survival capacity, stress tolerance, plant architecture  
CC or physiology, altered endoreduplication, biochemistry, signal  
CC transduction, storage lipid mobilization and/or altered photosynthesis,  
CC each relative to the corresponding wild type plants. Accordingly, these  
CC sequences can also be useful as positive or negative selectable markers  
CC during transformation of cells or tissues. The identified genes play a  
CC role in a variety of biological processes such as DNA replication, cell  
CC wall biosynthesis, nitrogen and/or carbon metabolism or they function as  
CC transcription factors. This polynucleotide sequence is thale cress cDNA  
CC repressed 1.3 fold or more in plants overexpressing the E2Fa/DPA  
CC transcription factor, given in an exemplification of the invention.

XX SQ Sequence 2286 BP; 535 A; 544 C; 462 G; 745 T; 0 U; 0 Other;

Query Match 73.6%; Score 18.4; DB 12; Length 2286;  
Best Local Similarity 95.0%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AACTCTCTGAGGCTTTTCAGT 24

||||| ||||| ||||| |||||

Db 1550 AACTCTCTGAGGCTTTTCAGT 1531

RESULT 31

ABN22936

ID ABN22936 standard; cDNA; 463 BP.

AC

ABN22936;

XX

24-JUN-2002 (first entry)

XX

Human ORFX polynucleotide sequence SEQ ID NO:14349.

XX

Human; open reading frame; ORFX; Gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis; gene; 88.

OS Homo sapiens.

XX

WO2001092523-A2.

XX

06-DEC-2001.

XX

29-MAY-2001; 2001WO-US010836.

XX

30-MAY-2000; 2000US-0206132P.

XX

29-AUG-2000; 2000US-0228716P.

XX

(CURA-) CURAGEN CORP.

XX

Shimkets RA, Leach MD;

XX

WPI; 2002-106308/14.

XX

P-FSDB; ABP07184.

XX

Novel human polypeptides and polynucleotides useful for diagnosing,

XX

preventing and treating cardiovascular disease, neurodegenerative,

XX

hyperproliferative disorders and autoimmune disorders.

XX

PS ...Disclosure; SEQ ID NO 14349; 1037pp; English.

XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIFO at ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 463 BP; 136 A; 100 C; 71 G; 156 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 6; Length 463;

Best Local Similarity 87.0%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGAGGCTTTTCAGT 24

||||| ||||| ||||| |||||

Db 198 AAAAATCTCTGAGGCTTTTCAGT 220

RESULT 32

ADH10017\_1

Continuation (2 of 4) of ADH10017 from base 100001 (Human chromosome 2p21-22 fragment  
WP Sequence split into 4 fragments LOCUS ADH10017 Accession Adh10017

WP Fragment Name Begin End

WP ADH10017\_0 1 110000

WP ADH10017\_1 100001 210000

WP ADH10017\_2 200001 310000

WP ADH10017\_3 300001 365186

Query Match 72.8%; Score 18.2; DB 10; Length 110000;

Best Local Similarity 87.0%; Pred. No. 4.9e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGAGGCTTTTCAGT 24

||||| ||||| ||||| |||||

Db 24978 AAAAATCTCTGAGGCTTTTCAGT 25000

RESULT 33

ADQ97266\_0

WP Sequence split into 4 fragments LOCUS ADQ97266 Accession Adq97266

WP Fragment Name Begin End

WP ADQ97266\_0 1 110000

WP ADQ97266\_1 100001 210000

WP ADQ97266\_2 200001 310000

WP ADQ97266\_3 300001 355211

ID ADQ97266 standard; DNA; 355211 BP.

XX

AC ADQ97266;

XX

07-OCT-2004 (first entry)

XX

Human cancer associated sequence HD08-023, SEQ ID 242.

XX

Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; da.

```
XX OS Homo sapiens.
XX PN WO2004060304-A2.
XX PD 22-JUL-2004.
XX PF 22-DEC-2003; 2003WO-US041389.
XX PR 27-DEC-2002; 2002US-00330773.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX DR WPI; 2004-543781/52.
XX PT New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX PS Claim 1; SEQ ID NO 242; 199pp; English.
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 355211 BP; 96742A; 68188C; 72417G; 106131T; 0U; 117330Other;

Query Match 72.8%; Score 18.2; DB 12; Length 110000;
Best Local Similarity 87.0%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 105281 AAAAAACACCTGTATGTTTTCAGT 105303

RESULT 34
ADQ97266_1
Continuation (2 of 4) of ADQ97266 from base 100001 (Human cancer associated sequence HDQ
WP Sequence split into 4 fragments LOCUS ADQ97266 Accession Adq97266
WP Fragment Name Begin End
WP ADQ97266_0 1 110000
WP ADQ97266_1 100001 210000
WP ADQ97266_2 200001 310000
WP ADQ97266_3 300001 355211

Query Match 72.8%; Score 18.2; DB 12; Length 110000;
Best Local Similarity 87.0%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 5281 AAAAAACACCTGTATGTTTTCAGT 5303

RESULT 35
ADQ97328_2/c
Continuation (3 of 4) of ADQ97328 from base 200001 (Mouse cancer associated sequence MDQ
WP Sequence split into 4 fragments LOCUS ADQ97328 Accession Adq97328
WP Fragment Name Begin End
WP ADQ97328_0 1 110000
WP ADQ97328_1 100001 210000
WP ADQ97328_2 200001 310000
WP ADQ97328_3 300001 358847

Query Match 72.8%; Score 18.2; DB 12; Length 110000;
Best Local Similarity 87.0%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 84160 AAAAAATCCTGTAGTCTTTTCAGT 84138

RESULT 36
ABD33491/c
ID ABD33491 standard; DNA; 112241 BP.
XX AC ABD33491;
XX DT 18-NOV-2004 (first entry)
XX DE Human cancer-associated (CA) gene HD07-095.
XX KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX KW ds; cancer; cytostatic.
XX OS Homo sapiens.
XX PN WO2004058146-A2.
XX PD 15-JUL-2004.
XX PF 15-DEC-2003; 2003WO-US040081.
XX PR 17-DEC-2002; 2002US-00322281.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX DR WPI; 2004-499109/47.
XX PT Novel human cancer associated protein encoded within open reading frame
XX PT of cancer associated gene, useful as targets for diagnosing cancer.
XX PS Claim 16; SEQ ID NO 656; 182pp; English.
XX CC The invention relates to cancer-associated proteins (CAP) and the cancer-
XX CC associated (CA) nucleic acids encoding them. The invention also relates
XX CC to a method for treating cancers involving administering to a patient an
XX CC inhibitor of CAP, and a method of screening for anticancer activity in a
XX CC potential drug involving providing a cell that expresses a CA gene
XX CC contacting a tissue sample derived from a cancer cell with an anticancer
XX CC drug candidate and monitoring the effect of the anticancer drug candidate
XX CC on expression of the CA gene. The CAP proteins are useful for detecting
XX CC cancer associated with expression of a CAP protein in a test cell sample
XX CC and for screening for a bioactive agent capable of modulating the
XX CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX CC cancer, involving determining the expression of a CA nucleic acid in a
XX CC tissue. This sequence represents a human CA gene of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 112241 BP; 28197 A; 21716 C; 21810 G; 31271 T; 0 U; 9247 Other;

Query Match 72.8%; Score 18.2; DB 13; Length 112241;
Best Local Similarity 87.0%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAGT 24
Db 84160 AAAAAATCCTGTGTTGCTTTTCAGT 84138

RESULT 37
ADL13676
ID ADL13676 standard; DNA; 192639 BP.
XX AC ADL13676;
```

XX 06-MAY-2004 (first entry)  
XX Osteoarthritis-associated polymorphic nucleotide #208.  
XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;  
XX joint space narrowing; osteophyte development; joint pain;  
XX osteoarthritis; SNP; single nucleotide polymorphism.  
XX Homo sapiens.  
XX WO2003054166-A2.  
XX 03-JUL-2003.  
XX 19-DEC-2002; 2002WO-US041225.  
XX 20-DEC-2001; 2001US-0342603P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Jones KA, Schafer A;  
XX WPI; 2003-559141/52.  
XX Determining susceptibility of an individual to joint space narrowing.  
XX osteophyte development and/or joint pain comprises identifying whether  
XX the individual has at least one polymorphism in a polynucleotide encoding  
XX a protein.  
XX Disclosure; SEQ ID NO 208; 297pp; English.  
XX The invention relates to a method of determining susceptibility of an  
XX individual to joint space narrowing and/or osteophyte development and/or  
XX joint pain comprising identifying whether the individual has at least one  
XX polymorphism in a polynucleotide encoding at least one of the protein  
XX listed in the specification. The methods, composition and agent are  
XX useful for modulating the susceptibility of an individual to joint space  
XX narrowing and/or osteophyte development and/or joint pain that is  
XX associated with a disease, preferably osteoarthritis. The cell line and  
XX the non-human animal are useful for screening for an agent for diagnosing  
XX an individual having susceptibility to joint space narrowing and/or  
XX osteophyte development and/or joint pain. This sequence corresponds to  
XX the polynucleotide encoding a protein listed in the specification. (Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences).  
XX SQ Sequence 192639 BP; 56569 A; 40174 C; 37841 G; 57918 T; 0 U; 137 Other;  
Query Match 72.8%; Score 18.2; DB 10; Length 192639;  
Best Local Similarity 87.0%; Pred. No. 5.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 3 AAAAATCTCTAGGCTTCAGTG 25  
Db 150038 AAAATCTCTAAGGCTATCAGTG 150060  
RESULT 38  
AAN82201/c  
ID AAN82201 standard; DNA; 1794 BP.  
XX AAN82201;  
XX 25-MAR-2003 (revised)  
XX 10-MAR-2003 (revised)  
XX 05-DEC-1990 (first entry)  
XX Beta-amylase from plant seed.  
XX Beta-amylase; plant seed; leader sequence; soybean; ss.

OS Glycine max.  
XX Key Location/Qualifiers  
FH 108..1595  
FT /\*tag= a  
FT /product= "beta-amylase"  
XX JP63091084-A.  
XX 21-APR-1988.  
XX 06-OCT-1986; 86JP-00236341.  
XX 06-OCT-1986; 86JP-00236341.  
XX (NORQ ) NORINSHO KK.  
XX WPI; 1988-150872/22.  
XX P-PSDB; AAP82687.  
XX Beta-amylase messenger RNA and novel recombinant - obtd. from plant seed  
XX by fractionation by saccharose density slope.  
XX Disclosure; Page ?; 6pp; Japanese.  
XX The messenger RNA corresp. to beta-amylase, obtd. from plant seed and  
XX free from leader sequence is obtd. in a fraction heavier than 28S by  
XX fractionation by saccharose density slope centrifugal sepn. method. Beta-  
XX amylase of soybean has heat resistance higher than beta-amylase from  
XX microorganisms. (Updated on 10-MAR-2003 to add missing OS field.)  
XX (Updated on 25-MAR-2003 to correct PA field.)  
XX SQ Sequence 1794 BP; 558 A; 372 C; 385 G; 479 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 1; Length 1794;  
Best Local Similarity 90.5%; Pred. NO. 4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AAAAACTCTGTAGGCTTTC 21  
Db 337 AACAACTCTGTAGGCTTTC 317  
RESULT 39  
ABQ69245\_03/c  
Continuation (4 of 31) of ABQ69245 from base 300001 (Listeria innocua DNA sequence #684  
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245  
WP Fragment Name Begin End  
WP ABQ69245\_00 1 110000  
WP ABQ69245\_01 100001 210000  
WP ABQ69245\_02 200001 310000  
WP ABQ69245\_03 300001 410000  
WP ABQ69245\_04 400001 510000  
WP ABQ69245\_05 500001 610000  
WP ABQ69245\_06 600001 710000  
WP ABQ69245\_07 700001 810000  
WP ABQ69245\_08 800001 910000  
WP ABQ69245\_09 900001 1010000  
WP ABQ69245\_10 1000001 1110000  
WP ABQ69245\_11 1100001 1210000  
WP ABQ69245\_12 1200001 1310000  
WP ABQ69245\_13 1300001 1410000  
WP ABQ69245\_14 1400001 1510000  
WP ABQ69245\_15 1500001 1610000  
WP ABQ69245\_16 1600001 1710000  
WP ABQ69245\_17 1700001 1810000  
WP ABQ69245\_18 1800001 1910000  
WP ABQ69245\_19 1900001 2010000  
WP ABQ69245\_20 2000001 2110000  
WP ABQ69245\_21 2100001 2210000  
WP ABQ69245\_22 2200001 2310000  
WP ABQ69245\_23 2300001 2410000  
WP ABQ69245\_24 2400001 2510000

```
WP ABQ69245_25 2500001 2610000
WP ABQ69245_26 2600001 2710000
WP ABQ69245_27 2700001 2810000
WP ABQ69245_28 2800001 2910000
WP ABQ69245_29 2900001 3010000
WP ABQ69245_30 3000001 3110000
WP ABQ69245_30 3000001 3110000

Query Match 71.2%; Score 17.8; DB 6; Length 110000;
Best Local Similarity 90.5%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAATCCTGTAGGCTTTCA 22
||||| ||||| |||||
Db 13500 AAAAATCCTGTAGGCTTTCA 13480

RESULT 40
ABQ67197_02/c
Continuation (3 of 12) of ABQ67197 from base 200001 (Listeria innocua contig DNA sequen
WP Sequence split into 12 fragments LOCUS ABQ67197 Accession Abq67197
WP Fragment Name Begin End
WP ABQ67197_00 1 110000
WP ABQ67197_01 100001 210000
WP ABQ67197_02 200001 310000
WP ABQ67197_03 300001 410000
WP ABQ67197_04 400001 510000
WP ABQ67197_05 500001 610000
WP ABQ67197_06 600001 710000
WP ABQ67197_07 700001 810000
WP ABQ67197_08 800001 910000
WP ABQ67197_09 900001 1010000
WP ABQ67197_10 1000001 1110000
WP ABQ67197_11 1100001 1163020

Query Match 71.2%; Score 17.8; DB 6; Length 110000;
Best Local Similarity 90.5%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAATCCTGTAGGCTTTCA 22
||||| ||||| |||||
Db 23235 AAAAATCCTGTAGGCTTTCA 23215

RESULT 41
AAL47655/c
ID AAL47655 standard; DNA; 43 BP.
XX AC AAL47655;
XX AC AAL47655;
DT 18-SEP-2002 (first entry)
XX DE Specific sequence isolation/determination method anchor primer #7.
XX DE Gene specific sequence isolation; sequence determination; disease;
XX KW diagnosis; infection; forensics; paternity testing; PCR; primer; anchor;
XX KW ss.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT modified_base 1 /*tag= a
FT FT /mod_base= OTHER
FT FT /note= "5' biotinylated"
XX PN WO200183696-A2.
XX PD 08-NOV-2001.
XX PF 27-APR-2001; 2001WO-US013807.
XX PR 28-APR-2000; 2000US-00560845.
XX XX
```

```
PA (DIGI-) DIGITAL GENE TECHNOLOGIES INC.
XX Muller R, Riddle GH, Glass JR;
XX WPI; 2002-114150/15.
XX Amplifying gene-specific(GS) polynucleotide by producing population of
PT DNA molecules having adapter molecule and anchor primer that has
PT captureable group, synthesizing and amplifying single-stranded GS
PT polynucleotide.
XX Claim 68; Page 102; 115pp; English.
XX The present invention relates to a method of isolating, amplifying and
XX sequencing gene specific sequences. This involves (a) synthesising double
XX stranded (ds) DNAs containing the gene specific polynucleotide which has
XX an anchor primer with a captureable group at the 3' end; (b) ligating a
XX ds adapter molecule to each dsDNA; (c) synthesising a single-stranded
XX (ss) gene specific polynucleotide using a gene specific oligonucleotide
XX primer having a second captureable group; and (d) amplifying the single
XX stranded gene specific polynucleotide after isolation. The method is used
XX for determining without cloning an extended sequence of a gene specific
XX polynucleotide that has a partial gene sequence known by carrying out the
XX method and then sequencing the amplified isolated gene specific
XX polynucleotide without cloning. The method has several diagnostic and
XX research applications, including genetic and infectious disease
XX diagnosis, forensic and paternity testing, toxicology testing, individual
XX identification, genetic research, occupational hazard screening and
XX pharmaceutical development. The method can be used to detect mutations in
XX target genes which may lead to disease conditions such as Alzheimer's
XX disease, cancer and Duchenne muscular dystrophy, thus facilitating the
XX diagnosis of genetic diseases as well as the identification of carriers
XX of recessive genetic disorders. The present sequence is an
XX oligonucleotide useful in the method of the invention
XX SQ Sequence 43 BP; 10 A; 4 C; 7 G; 22 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 43;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAATCCTGTAGGCTTTCA 24
||||| ||||| |||||
Db 31 AAAAATCCTGTAGGCTTTCA 8

RESULT 42
AAL47656/c
ID AAL47656 standard; DNA; 43 BP.
XX AC AAL47656;
XX AC AAL47656;
DT 18-SEP-2002 (first entry)
XX DE Specific sequence isolation/determination method anchor primer #8.
XX DE Gene specific sequence isolation; sequence determination; disease;
XX KW diagnosis; infection; forensics; paternity testing; PCR; primer; anchor;
XX KW ss.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT modified_base 1 /*tag= a
FT FT /mod_base= OTHER
FT FT /note= "5' aminated"
XX PN WO200183696-A2.
XX PD 08-NOV-2001.
XX PF 27-APR-2001; 2001WO-US013807.
XX XX
```



```
XX 28-APR-2000; 2000US-00560845.
PR (DIGI-) DIGITAL GENE TECHNOLOGIES INC.
XX Muller R, Riddle GH, Glass JR;
XX WPI; 2002-114150/15.
XX Amplifying gene-specific(GS) polynucleotide by producing population of
PT DNA molecules having adapter molecule and anchor primer that has
PT captureable group, synthesizing and amplifying single-stranded GS
PT polynucleotide.
XX Claim 68; Page 103; 115pp; English.
XX The present invention relates to a method of isolating, amplifying and
CC sequencing gene specific sequences. This involves (a) synthesizing double
CC -stranded (ds) DNAs containing the gene specific polynucleotide which has
CC an anchor primer with a captureable group at the 3' end; (b) ligating a
CC ds adapter molecule to each dsDNA; (c) synthesizing a single-stranded
CC (ss) gene specific polynucleotide using a gene specific oligonucleotide
CC primer having a second captureable group; and (d) amplifying the single
CC stranded gene specific polynucleotide after isolation. The method is used
CC for determining without cloning an extended sequence of a gene specific
CC polynucleotide that has a partial gene sequence known by carrying out the
CC method and then sequencing the amplified isolated gene specific
CC polynucleotide without cloning. The method has several diagnostic and
CC research applications, including genetic and infectious disease
CC diagnosis, forensic and paternity testing, toxicology testing, individual
CC identification, genetic research, occupational hazard screening and
CC pharmaceutical development. The method can be used to detect mutations in
CC target genes which may lead to disease conditions such as Alzheimer's
CC disease, cancer and Duchenne muscular dystrophy, thus facilitating the
CC diagnosis of genetic diseases as well as the identification of carriers
CC of recessive genetic disorders. The present sequence is an
CC oligonucleotide useful in the method of the invention
XX
SQ Sequence 43 BP; 10 A; 4 C; 7 G; 22 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 6; Length 43;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAATCCTGTAGGCTTTTCAGT 24
Db 31 AAAAAAATCCTGTAGGCTTTTCAGT 8
RESULT 43
ADB50438/c
ID ADB50438 standard; DNA; 457 BP.
XX ADB50438;
XX 04-DEC-2003 (first entry)
XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:980.
XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
XX toxicity marker; toxicity progression; drug screening;
XX primary rat hepatocyte toxicity modelling; gene; ds.
XX Rattus norvegicus.
XX WO2003065993-A2.
XX 14-AUG-2003.
XX 04-FEB-2003; 2003WO-US003482.
XX 04-FEB-2002; 2002US-0353171P.
XX 13-MAR-2002; 2002US-0363534P.
```

```
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX (GENE-) GENE LOGIC INC.
PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
XX Blashoff M;
PI WPI; 2003-731472/69.
XX Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX Claim 44; SEQ ID NO 980; 874pp; English.
XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
SQ Sequence 457 BP; 139 A; 103 C; 102 G; 113 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 10; Length 457;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAATCCTGTAGGCTTTTCAGT 24
Db 262 AAAAAAATCCTGTAGGCTTTTCAGT 239
RESULT 44
ADB56904/c
ID ADB56904 standard; cDNA; 925 BP.
XX ADB56904;
XX 02-DEC-2004 (first entry)
XX Bacterial polynucleotide #8891.
XX Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
```

KW bacterial polynucleotide; gene; ss.  
 XX Bacteria.  
 OS US2003233675-A1.  
 XX 18-DEC-2003.  
 XX 20-FEB-2003; 2003US-00369493.  
 XX 21-FEB-2002; 2002US-0360039P.  
 XX (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 PI WPI; 2004-061375/06.  
 XX New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX Claim 1; SEQ ID NO 32578; 122pp; English.  
 XX The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at seqdata.uspto.gov/sequence.html.  
 XX Sequence 925 BP; 168 A; 237 C; 278 G; 242 T; 0 U; 0 Other;  
 SQ Query Match 70.4%; Score 17.6; DB 13; Length 925;  
 Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 AAAAAGCTCTGAGGCTTTTCAGTG 25  
 Db 912 AAAAAGCTCTGAGGCTTTTCAGTG 889  
 RESULT 45  
 ADX09691/C  
 ID ADX09691 standard; cDNA; 1303 BP.  
 XX AC ADX09691;  
 XX 21-APR-2005 (first entry)  
 XX Plant full length insert polynucleotide seqid 4266.  
 DE

XX plant protectant; plant growth regulant; gene therapy; plant;  
 KW recombinant DNA construct; physical array; plant breeding marker;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
 KW growth rate; cell cycle pathway; disease resistance;  
 KW galactomannan production; lignin production; plant growth regulator;  
 KW yield; plant growth; plant development; seed oil; protein yield;  
 KW protein content; gene; ss.  
 XX Unidentified.  
 OS US2004034888-A1.  
 XX 19-FEB-2004.  
 XX 28-APR-2003; 2003US-00425114.  
 XX 06-MAY-1999; 99US-00304517.  
 XX 05-NOV-2001; 2001US-00985678.  
 XX (LIUJ/) LIU J.  
 PA (ZHOU/) ZHOU Y.  
 PA (KOVA/) KOVALIC D K.  
 PA (SCRE/) SCREEN S E.  
 PA (TABA/) TABASKA J E.  
 PA (CAOY/) CAO Y.  
 XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
 PI WPI; 2004-180133/17.  
 XX New recombinant DNA construct, useful for improving plant tolerance to  
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
 PT pests, for conferring increased resistance to plant disease, or for  
 PT improving yield.  
 XX Claim 1; SEQ ID NO 4266; 15pp; English.  
 XX The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at  
 CC ftp.seqdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC increased resistance to plant disease, for producing galactomannan,  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This sequence represents a plant full length insert  
 CC polynucleotide that can be used in the recombinant DNA construct of the  
 CC invention.  
 XX Sequence 1303 BP; 381 A; 278 C; 257 G; 387 T; 0 U; 0 Other;  
 SQ Query Match 70.4%; Score 17.6; DB 13; Length 1303;  
 Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAGCTCTGAGGCTTTTCAGT 24  
 Db 208 AAAAAGCTCTGAGGCTTTTCAGT 185  
 RESULT 46  
 ABL11811  
 ID ABL11811 standard; cDNA; 1368 BP.  
 XX

AC ABL11811;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29915.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical; gene; ss.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WIPI; 2001-656860/75.  
 XX P-PSDB; ABB67708.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX Claim 1; SEQ ID NO 29915; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1368 BP; 329 A; 347 C; 359 G; 333 T; 0 U; 0 Other;  
 SQ Query Match 70.4%; Score 17.6; DB 4; Length 1368;  
 Best Local Similarity 83.3%; Pred. No. 4.9e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAAAGCTCTGTAGGCTTTTCAGT 24  
 DB 534 AGAAAGCTCTGTAGGCTTTTCAGT 557  
 RESULT 47  
 AAQ99349  
 ID AAQ99349 standard; cDNA; 1563 BP.  
 XX AC AAQ99349;  
 XX DT 24-APR-1996 (first entry)  
 XX DE Frog bombesin-like peptide receptor BB4 cDNA.  
 XX KW Bombesin-like peptide receptor BB4; small cell lung cancer; diagnosis;  
 XX KW therapy; ss.  
 XX OS Bombina orientalis.  
 XX PN Key Location/Qualifiers  
 XX FT 192..1322  
 XX FT /\*tag= a

XX WO9603416-A1.  
 XX PD 08-FEB-1996.  
 XX PF 21-JUL-1995; 95WO-US009157.  
 XX PR 22-JUL-1994; 94US-00279590.  
 XX PA (OREG-) OREGON REGIONAL PRIMATE RES CENT.  
 XX PI Spindel ER, Nagalla S, Barry BJ;  
 XX WIPI; 1996-116984/12.  
 XX P-PSDB; AAR89359.  
 XX Nucleic acid encoding bombesin-like peptide receptor - used to screen for  
 PT cpds. which interact with receptor, isolated cpds. can then be used in  
 PT diagnosis and treatment of, e.g. human small cell lung cancer.  
 XX Claim 1; Page 15-17; 33pp; English.  
 XX cDNA (AAQ99349) coding for the frog (Bombina orientalis) novel bombesin-  
 CC like peptide receptor BB4 (AAR89359) was obt'd. from a brain cDNA library  
 CC in vector lambda ZAP II using a cRNA probe derived from an isolated  
 CC sequence of the receptor. The cDNA was used to probe a monkey brain cDNA  
 CC library, identifying 2 partial coding sequences (AAQ99350 and AAQ99351)  
 CC for monkey BB4 (AAR89360) and BB5 (AAR89361). The cDNA can be  
 CC incorporated into a vector and expressed in host cells to provide pure  
 CC BB4  
 XX Sequence 1563 BP; 453 A; 329 C; 298 G; 483 T; 0 U; 0 Other;  
 SQ Query Match 70.4%; Score 17.6; DB 2; Length 1563;  
 Best Local Similarity 83.3%; Pred. No. 4.9e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAAAGCTCTGTAGGCTTTTCAGT 24  
 DB 187 AAAAAAGCTCTGAAGGTTTTCAGT 210  
 RESULT 48  
 AAS12704  
 ID AAS12704 standard; cDNA; 1563 BP.  
 XX AC AAS12704;  
 XX DT 04-DEC-2001 (first entry)  
 XX DE Frog bombesin-like peptide BB4 receptor cDNA.  
 XX KW Bombesin-like peptide receptor; vertebrate; small cell lung cancer; ss;  
 XX KW gastrin releasing peptide; neuromedin B; neuromedin C; testis; uterus;  
 XX KW bombesin receptor subtype-3; BRS-3; gastrointestinal tract; NMB; NMC;  
 XX KW central nervous system; BB4 receptor; BB5 receptor; GRP; frog.  
 XX OS Bombina orientalis.  
 XX PN Key Location/Qualifiers  
 XX FT 192..1322  
 XX FT /\*tag= a  
 XX FT /product= "Frog BB4 receptor"  
 XX US2001014457-A1.  
 XX PN 16-AUG-2001.  
 XX PD 24-SEP-1998; 98US-00160116.  
 XX PF 22-JUL-1994; 94US-00279590.  
 XX PR 12-AUG-1997; 97US-00910092.  
 XX

PA (SPIN/) SPINDEL E R.  
PA (NAGA/) NAGALLA S.  
PA (BARR/) BARRY B.  
XX  
PI Spindel ER, Nagalla S, Barry B;  
XX  
DR WPI: 2001-488385/53.  
DR P-PSDB; AAU07635.  
XX  
PT New bombesin like peptide with high binding affinity for bombesin  
PT receptors for the modulation of receptors, for example, in the  
PT gastrointestinal tract.  
XX  
PS Claim 15; Fig 1; 16pp; English.  
XX  
CC The invention relates to genes which encode receptors capable of binding  
CC to bombesin-like peptides, expressed in vertebrate tissue. Bombesin-like  
CC peptides are produced and secreted by human small cell lung cancer cells.  
CC Examples of these peptides are gastrin releasing peptide (GRP),  
CC neuromedin B (NMB), neuromedin C (NMC), bombesin receptor subtype-3 (BRS-  
CC 3), BB4 and BB5. The peptides of the invention are useful for modulating  
CC receptors to which they have an affinity including receptors in the  
CC gastrointestinal tract, the central nervous system, testes and uterus.  
CC The sequences are therefore used in the treatment of cancers such as  
CC small cell lung cancer. This sequence represents a cDNA encoding the frog  
CC BB4 receptor  
XX  
SQ Sequence 1563 BP; 450 A; 333 C; 296 G; 484 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 5; Length 1563;  
Best Local Similarity 83.3%; Pred. No. 4.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTCCTGCTAGGCTTTCAGT 24  
DB 187 AAAAAATGCTGAAGGTTTTCAGT 210  
RESULT 49  
ADE83485/c  
ID ADE83485 standard; DNA; 2962 BP.  
XX  
AC ADE83485;  
XX  
XX 29-JAN-2004 (first entry)  
XX Human gene NM\_005419, SEQ ID NO 11081.  
XX  
XX Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
XX Homo sapiens.  
XX WO2003016475-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI: 2003-268312/26.  
DR GENBANK; NM\_005419.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2962 BP; 749 A; 763 C; 758 G; 692 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 10; Length 2962;  
Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTCCTGCTAGGCTTTCAGT 24  
DB 520 AGAGAGACTCCTGCTGCTTTCAGT 497  
RESULT 50  
ABLL1810/c  
ID ABL11810 standard; cDNA; 5010 BP.  
XX  
AC ABL11810;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 29912.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR

DR P-PSDB; ABB67707.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX  
PS Claim 1; SEQ ID NO 29912; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5010 BP; 1462 A; 1029 C; 1048 G; 1471 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 4; Length 5010;  
Best Local Similarity 83.3%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTCCTGTAGGCTTTCAGT 24  
Db 1957 AGAAAACTCCTGTGGGCTATCAGT 1934

Search completed: February 3, 2006, 21:56:58  
Job time : 212.111 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-20  
Perfect score: 25  
Sequence: 1 aaaaactcctgtaggcttctcagt 25  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_gse1.\*  
10: gb\_gse2.\*  
11: gb\_gse3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	706	2	BB474887 BB474887
2	25	100.0	3793	4	AK084510 Mus muscu
3	25	100.0	4437	4	AK036338 Mus muscu
4	20.2	80.8	697	2	B218988 601882209
5	19.8	79.2	634	8	DN755785 GL-Cf-125
6	19.8	79.2	704	8	DN267760 L1B30320
7	19.8	79.2	715	8	DN265405 L1B30320
8	19.8	79.2	765	8	DN392910 L1B3934-0
9	19.8	79.2	793	8	DN370580 L1B3733-0
10	19.8	79.2	892	8	CV804204 AGENCOURT
11	19.8	79.2	970	6	CD502134 CDA55-A01
12	19.4	77.6	750	5	BX921863 BX921863
13	19.4	77.6	961	11	CNS04CQT AL284798 Tetraodon
14	19.2	76.8	154	1	A1918165 nt08a05.x
15	19.2	76.8	163	6	CD544491 B0258C10-
16	19.2	76.8	169	3	BM769029 K-EST0052
17	19.2	76.8	186	1	AA864560 oh31b01.s
18	19.2	76.8	203	1	AI886941 w194C12.x
19	19.2	76.8	226	1	AV232251 AV232251
20	19.2	76.8	226	3	BI687636 603312626
21	19.2	76.8	236	1	AV255052 AV255052
22	19.2	76.8	239	5	BY372216 BY372216



1	(bases 1 to 706)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
---	------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

pBluescript KS(+) after bulk excision from Lambda FLC I."

```

ORIGIN
Query Match      100.0%; Score 25; DB 2; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGTG 25
    |||||
Db 575 AAAAAAATCTCTAGGCTTTTCAGTG 599

RESULT 2
AK084510
LOCUS
DEFINITION
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
library, clone:D330010D18 product:mesoderm development candidate 1,
full insert sequence.

ACCESSION
AK084510.1 GI:26351134
VERSION
AK084510.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
AUTHORS
Carninci, P., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
JOURNAL
PUBMED
REFERENCE
2
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
JOURNAL
PUBMED
REFERENCE
3
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, O.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
4
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
5
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
6
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sgobe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

```

```

TITLE
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
Location/Qualifiers
1..3793
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:D330010D18"
/db_xref="taxon:10090"
/clone="D330010D18"
/tissue_type="heart"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
1067..2155
/note="unannotated protein product; mesoderm development
candidate 1 (MGD|GRI:1891420, GB|NM_030705, evidence:
BLASTN, 99%, match=3258)
putative"
/codon_start=1
/protein_id="BAC39204.1"
/db_xref="GI:26351135"
/translation="MASGAGKPTGEASPAGSAPGSSQPRKRLVSCDHCKGKM
QLVADLLLSSEARPVLFEGPASPGAGAESFEQRTDIIARTKGLSLTHDVOSQLNM
GFGEGADSLVELDVLVSLTBCSAHAAYLAAVATPAQAPQGLVDRYVTRCRHEV
EGCAVLATPLADMTFQLLLEVSQGLSRNLKELTDACALASDKSRDRFREFQKLG
KCMZSALLACREVKAAPSRKALRSCALFSGPLVQAVSLVGFATEPQFQIGRAA
VTEGKAVQTALGGAMSVSACVLLTQCLRLAQHPDGSKMSDHRRLRNSACAVS
EGCTLLSQALRRSSPRTLPPVNSNVN"
ORIGIN
Query Match      100.0%; Score 25; DB 4; Length 3793;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGTG 25
    |||||
Db 3662 AAAAAAATCTCTAGGCTTTTCAGTG 3686

RESULT 3
AK036338
LOCUS
DEFINITION
Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:9630059C21 product:mesoderm development
candidate 1, full insert sequence.

ACCESSION
AK036338.1 GI:26085172
VERSION
AK036338.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
AUTHORS
Carninci, P., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
JOURNAL
PUBMED
REFERENCE
2
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
JOURNAL
PUBMED
REFERENCE
3
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, O.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
4
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
5
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
PUBMED
11076861
JOURNAL
PUBMED
REFERENCE
6
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sgobe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

```

```
2
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE
AUTHORS 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
PUBMED 11076861
REFERENCE
AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS 5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS 6 (bases 1 to 4437)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
Location/Qualifiers
1..4437
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="FANTOM DB:9630059C21"
/db_xref="taxon:10090"
/clone="9630059C21"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days neonate"
1086..2175
/notes="mesoderm development candidate 1 (MGD|MG1:1891420,
GB|NM_030705, evidence: BLASTN, 99%, match=3258)"

misc_feature
557 AAAAATTCCTGTAGGTTTCATTG 533

Query Match 80.8%; Score 20.2; DB 2; Length 697;
Best Local Similarity 88.0%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
Query 1 AAAAATTCCTGTAGGTTTCATTG 25
+++++
557 AAAAATTCCTGTAGGTTTCATTG 533
+++++

polyA_signal putative"
4413..4418
/notes="putative"
polyA_site 4437
/notes="putative"

ORIGIN
Query Match 100.0%; Score 25; DB 4; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAATTCCTGTAGGTTTCATTG 25
+++++
3682 AAAAATTCCTGTAGGTTTCATTG 3706
+++++

Db 3682 AAAAATTCCTGTAGGTTTCATTG 3706
+++++

RESULT 4
BF218988/c 697 bp mRNA linear EST 06-NOV-2000
LOCUS 601882209F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4094499 5',
DEFINITION mRNA sequence.
ACCESSION BF218988
VERSION BF218988.1 GI:11112575
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS NIH-MGC http://mgc.mci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW955 row: b column: 04
High quality sequence stop: 613.
FEATURES
source
Location/Qualifiers
1..697
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4094499"
/tissue_type="glioblastoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 57"
/notes="Organ: brain; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggcgcctcgcc); Site 2: SfiI (ggcattatggcc);
SfiI (ggcgcctcgcc); Site 2: SfiI (ggcattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATATGACC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGAGCGGCGCATG-dt(30)BN-3'
(where B = A, C, G or T and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
```



```

JOURNAL      Unpublished (2005)
COMMENT      Contact: Nick Staten
             Tel: 636 247 6855
             Email: nicholas.r.staten@pfizer.com.

FEATURES
source
1..765
/organism="Canis familiaris"
/mol_type="mRNA"
/strain="beagle"
/db_xref="taxon:9615"
/clone="CLN12935017"
/tissue_type="brain"
/lab_host="DH108"
/clone_lib="LIB3934"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; brain
from normal dog"

ORIGIN
Query Match          79.2%; Score 19.8; DB 8; Length 765;
Best Local Similarity 91.3%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  AAAAATCCTCTGAGGCTTTCAGT 24
        ||||||| ||||||| ||||||| |||||||
Db      375 AAAAATCCTCTGAGGCTATCAGT 397

RESULT 9
DN370580      793 bp      mRNA      linear      EST 07-MAR-2005
LOCUS      LIB3733-010-A1-K1-A12 LIB3733 Canis familiaris cDNA clone
DEFINITION      CLN12917873, mRNA sequence.
ACCESSION      DN370580
VERSION      DN370580.1 GI:60550723
KEYWORDS      EST.
SOURCE      Canis familiaris (dog)
ORGANISM      Canis familiaris
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
             Canis.
REFERENCE      1. (bases 1 to 793)
AUTHORS      Staten,N.R.
TITLE      Direct Submission (Staten,N.R.)
JOURNAL      Unpublished (2005)
COMMENT      Contact: Nick Staten
             Tel: 636 247 6855
             Email: nicholas.r.staten@pfizer.com.

FEATURES
source
1..793
/organism="Canis familiaris"
/mol_type="mRNA"
/strain="beagle"
/db_xref="taxon:9615"
/clone="CLN12917873"
/tissue_type="brain"
/lab_host="BHK"
/clone_lib="LIB3733"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; brain
-dog"

ORIGIN
Query Match          79.2%; Score 19.8; DB 8; Length 793;
Best Local Similarity 91.3%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  AAAAATCCTCTGAGGCTTTCAGT 24
        ||||||| ||||||| ||||||| |||||||
Db      642 AAAAATCCTCTGAGGCTATCAGT 664

RESULT 10
CV804204      892 bp      mRNA      linear      EST 21-JAN-2005
LOCUS

```

---

```

DEFINITION      AGENCOURT 36106332 NIH XGC tropKid1 Xenopus tropicalis cDNA clone
IMAGE:7461930 5', mRNA sequence.
ACCESSION      CV804204
VERSION      CV804204.1 GI:55747170
KEYWORDS      EST.
SOURCE      Xenopus tropicalis (western clawed frog)
ORGANISM      Xenopus tropicalis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
             Xenopodinae; Xenopus; Silurana.
REFERENCE      1 (bases 1 to 892)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Daniela S. Gerhard, Ph.D.
             Office of Cancer Genomics
             National Cancer Institute / NIH
             Bldg. 31 Rm10A07 Bethesda, MD 20892
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Bruce Blumberg
             cDNA Library Preparation: B. Blumberg
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM15758 row: d column: 16
             High quality sequence start: 10
             High quality sequence stop: 672.
             Location/Qualifiers
               1..892
                 /organism="Xenopus tropicalis"
                 /mol_type="mRNA"
                 /strain="N6 (Nigerian 6th generation inbred)"
                 /db_xref="taxon:8364"
                 /clone="IMAGE:7461930"
                 /tissue_type="Kidney"
                 /dev_stage="Adult"
                 /lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
                 /clone_lib="NIH_XGC_tropKid1"
                 /note="Vector: pCS107; Site_1: EcoRI; Site_2: XhoI; The
                 library was prepared from 5 ug of poly A+ RNA by oligo-dT
                 priming
                 [5'-GAGAGAGAGAGAGAGACTAGTCTCGAGTGTCTTTTCTTTT-3']
                 and Stratascript reverse transcriptase. After ligation of
                 EcoRI adapters (5'-AATTCGACGAGG-3') followed by kinasing
                 adapters and by XhoI digestion, the cDNA was size selected
                 by chromatography on Sepharose CL-2B columns and fractions
                 containing cDNAs larger than 1000 bp were ligated into
                 EcoRI/XhoI-digested pCS107. Reference for library
                 construction: Current Genomics 4, 635-644. Library
                 constructed by Michelle Tabb and Bruce Blumberg (Dept of
                 Developmental and Cell Biology, University of California,
                 Irvine).".
ORIGIN
Query Match          79.2%; Score 19.8; DB 8; Length 892;
Best Local Similarity 91.3%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  AAAAATCCTCTGAGGCTTTCAGT 24
        ||||||| ||||||| ||||||| |||||||
Db      342 AAAAATCCTCTGAGGCTGTCACT 364

RESULT 11
CD502134
LOCUS      CD502134
DEFINITION      CD455-A01.yld-s SHGC-CDA Gasterosteus aculeatus cDNA clone
ACCESSION      CD502134
VERSION      CD502134.1 GI:31429160
KEYWORDS      EST.

```

**SOURCE ORGANISM**  
 Gasterosteus aculeatus (three spined stickleback)  
 Gasterosteus aculeatus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
 Gasterosteidae; Gasterosteus.  
 1 (bases 1 to 970)

**REFERENCE AUTHORS**  
 Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,  
 Schmutz, J. and Myers, R.M.  
 Expressed sequence tags from Gasterosteus aculeatus  
 Unpublished (2003)  
 Contact: Kingsley, DM  
 HHMI and Department of Developmental Biology  
 Stanford University School of Medicine  
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA  
 Tel: 650 725 5954  
 Fax: 650 725 7739  
 Email: kingsley@cngm.stanford.edu  
 Plate: 55  
 High quality sequence start: 5  
 High quality sequence stop: 837.  
 Location/Qualifiers  
 1..970

**FEATURES**  
 source  
 /organism="Gasterosteus aculeatus"  
 /mol\_type="mRNA"  
 /strain="Salinas river, CA"  
 /db\_xref="taxon:69293"  
 /clone="CD45-A01"  
 /sex="mixed male and female"  
 /tissue\_type="heads and internal organs combined"  
 /dev\_stage="adult"  
 /clone\_lib="SHGC-CDA"  
 /note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI  
 (5' adaptor); Site 2: XhoI (3' linker primer); The mixed  
 organ cDNA library was generated using the ZAP-cDNA method  
 by Stratagene. First strand cDNA synthesis was primed with  
 a 50 bp linker primer containing an oligo dt sequence  
 preceded by a synthetic XhoI site. 5 prime adaptors were  
 used containing an EcoRI cohesive end. The finished cDNAs  
 were inserted in to the ZAP express vector  
 unidirectionally in the sense orientation with respect to  
 the lacZ promoter of pBK-CMV. An amplified library was  
 prepared from approximately 3 million primary clones in  
 the lambda ZAP Express vector. In vivo excision was then  
 used to generate individual pBK-CMV phagemid clones for  
 EST sequencing."

**ORIGIN**  
 Query Match 79.2%; Score 19.8; DB 6; Length 970;  
 Best Local Similarity 91.3%; Pred. No. 3.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAAAACTCTGTAGGCTTCAGT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 38 AAAACACCTGGAGGCTTCAGT 60

**RESULT 12**  
 BX921863/c  
 LOCUS  
 DEFINITION  
 BX921863 Sus Scrofa library (scan) Sus scrofa cDNA clone  
 scan0019d.o.08 5prim, mRNA sequence.  
 BX921863  
 EST.  
 BX921863.1 GI:41138711  
 EST.  
 Sus scrofa (pig)  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
 Sus.  
 1 (bases 1 to 750)  
 Bonnet, A., Tosser-Klopp, G., Benne, F., Cabau, C., Villegier, S.,  
 Soares, M., Bonaldo, F. and Hatey, F.

**SOURCE ORGANISM**  
 A Pig Normalised Multi-Tissue cDNA Library  
 Unpublished (2003)  
 Contact: Tosser-Klopp G  
 Genetique Animale  
 Institut National de la Recherche Agronomique  
 Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan  
 cedex, FRANCE  
 Tel: 33 (0) 5.61.28.51.14  
 Fax: 33 (0) 5.61.28.53.08  
 Email: tosser@toulouse.inra.fr  
 Sequence cleaned of vector, adaptor and repetitions. Contact us  
 at aigenasupport@jouy.inra.fr to obtain the Chromatogram of this  
 sequence.  
 Plate: 0019 row: 0 column: 8.  
 Location/Qualifiers  
 1..750  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /clone="scan0019d.o.08"  
 /tissue\_type="mixed"  
 /dev\_stage="from embryos to adults"  
 /clone\_lib="Sus Scrofa library (scan)"  
 /note="tissues: adipose tissue, brain, kidney, liver,  
 muscle, ovary, testis, heart, hypothalamus, pancreas,  
 skin, spleen, thymus, placenta, pituitary gland, seminal  
 vesicle, small intestine, uterus, adrenals, bulbo urethral  
 gland, cerebral trunk, epididymis, female gonad,  
 gall bladder, hippocampus, large intestine, male gonad,  
 melanocytes, stomach, udder"

**ORIGIN**  
 Query Match 77.6%; Score 19.4; DB 5; Length 750;  
 Best Local Similarity 95.2%; Pred. No. 5.5e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAACTCTGTAGGCTTTC 21  
 ||||| ||||| ||||| ||||| |||||  
 Db 412 AAAAAGTCCTGTAGGCTTTC 392

**RESULT 13**  
 CNS04CQT/c  
 LOCUS  
 DEFINITION  
 CNS04CQT 961 bp DNA linear GSS 01-SEP-2000  
 Tetraodon nigroviridis genome survey sequence T7 end of clone  
 101C23 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 AL284798  
 AL284798.1 GI:8023197  
 GSS; genome survey sequence.  
 Tetraodon nigroviridis  
 Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontoidea; Tetraodontidae; Tetraodon.  
 1  
 Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
 Saurin, W. and Weissenbach, J.  
 Estimate of human gene number provided by genome-wide analysis  
 using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)  
 10835645  
 2  
 Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,  
 Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
 Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)  
 10899143  
 3 (bases 1 to 961)

**REFERENCE AUTHORS**  
 Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,  
 Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
 Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)  
 10899143  
 3 (bases 1 to 961)



AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
FEATURES  
source  
Location/Qualifiers  
1..961  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="101C23"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COBG101AB12LP1  
end : T7"

ORIGIN  
Query Match 77.6%; Score 19.4; DB 11; Length 961;  
Best Local Similarity 95.2%; Pred. No. 5.8e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 AAAACTCCGTAGGCTTTCAG 23  
|||||  
Db 462 AAAACTCCGTAGGCTTTCG 442

RESULT 14  
AI918165  
LOCUS  
DEFINITION tn08a05.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2166992 3',  
mRNA sequence.  
ACCESSION AI918165  
VERSION AI918165.1 GI:5638020  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 154)  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1243 Std Error: 0.00  
Seq primer: -400P from Gbco  
High quality sequence stop: 134.  
Location/Qualifiers  
1..154  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2166992"  
/issue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Brn25"  
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTCGAGCGCCGCATAGGTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 76.8%; Score 19.2; DB 1; Length 154;  
Best Local Similarity 87.5%; Pred. No. 5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAATCTCTGTAGGCTTTCAGT 24  
|||||  
Db 130 AAAAATCTCTGTAGGCTTTCATT 153

RESULT 15  
CD544491  
LOCUS  
DEFINITION B0258C10-3 NTA Mouse Embryonic Germ Cell cDNA Library (Long) Mus  
musculus cDNA clone NIA:B0258C10 IMAGE:30108993 3', mRNA sequence.  
ACCESSION CD544491  
VERSION CD544491.1 GI:31592226  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 163)  
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.  
Construction of long-transcript enriched cDNA libraries from  
submicrogram amounts of total RNAs by a universal PCR amplification  
method  
Genome Res. 11 (9), 1553-1558 (2001)  
11544199  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: B0258 row: C column: 10  
Seq primer: -21M13 Forward  
High quality sequence stop: 163  
POLYA=Yes.  
Location/Qualifiers  
1..163  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="niaEST:B0258C10-3"  
/db\_xref="taxon:10090"  
/clone="NIA:B0258C10 IMAGE:30108993"  
/sex="male"  
/dev\_stage="embryonic day 8"  
/lab\_host="DH10B"  
/clone\_lib="NTA Mouse Embryonic Germ Cell cDNA Library  
(Long)"  
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).  
This is a long-transcript enriched cDNA library (Ref.  
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total  
RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP).  
EG cells were cultured at 37 C, 5% CO2 in DMEM  
supplemented with 15% ES cell-qualified FBS, 0.1mM  
non-essential amino acids, 2 mM glutamine,  
penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM  
beta-mercaptoethanol, and 1000000 units of LIF per liter.  
Double-stranded cDNAs were synthesized with an Oligo(dT)



primer [Invitrogen:  
5'-pGACTAGTCTAGATCGGAGCGGCCCGCTTTT-3'] from  
2.5 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to lona-linker Lu-Sal4, purified by Centricon 100. Then, the  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 4.0 kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 163;  
Best Local Similarity 87.5%; Pred. No. 5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
Db 123 AAAAAATTCCTGTAGGTTTCATT 146

RESULT 16  
BM769029/c  
LOCUS  
DEFINITION BM769029 169 bp mRNA linear EST 04-MAR-2002  
K-EST0052192 S14K402 Homo sapiens cDNA clone S14K402-14-B08 5',  
mRNA sequence.

ACCESSION BM769029.1 GI:19098644  
VERSION  
KEYWORDS EST.

## ORGANISM

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 169)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

## JOURNAL

COMMENT Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 14 row: B column: 08

High quality sequence stop: 169.

Location/Qualifiers

## FEATURES

## source

1..169  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S14K402-14-B08"  
/cell\_line="K402"  
/lab\_host="Top10F"  
/clone\_lib="S14K402"  
/note="Organ: Stomach; Vector: pTZ19RP1; Site:1: EcoRI;  
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 169;  
Best Local Similarity 87.5%; Pred. No. 5.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
Db 36 AAAAAATTCCTGTAGGTTTCATT 13

## RESULT 17

AA864560

LOCUS

DEFINITION AA864560 186 bp mRNA linear EST 13-MAY-1998  
oh31b01.s1 NCI\_CGAP\_Kid6 Homo sapiens cDNA clone IMAGE:1459369 3',  
mRNA sequence.

ACCESSION AA864560.1 GI:2958873

VERSION

KEYWORDS EST.

SOURCE

## ORGANISM

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 186)

## AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## TITLE

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Cloned through the I.M.A.G.E. Consortium/LENL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert length: 1284 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 111.

Location/Qualifiers

1..186

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1459369"

/sex="mixed"

/tissue\_type="kidney tumor"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="NCI CGAP\_Kid6"

/note="Organ: kidney; Vector: Bluescript SK-; Site:1:

EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pooled kidney tumors. 5' adaptor sequence: 5'

GAATTCGGCACGAG 3' 3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 186;  
Best Local Similarity 87.5%; Pred. No. 5.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
Db 145 AAAAAATTCCTGTAGGTTTCATT 168

RESULT 18	ORGANISM	Mus musculus
AI886941		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
LOCUS		
DEFINITION		
AI886941		1 (bases 1 to 226)
ACCESSION		Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
REFERENCE		RIKEN Mouse ESTs (Konno,H., et al. 1999)
AUTHORS		Unpublished (1999)
TITLE		Contact: Yoshihide Hayashizaki
JOURNAL		Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
COMMENT		The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
		Saeaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
		Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
		Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
		Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
		Carninci,P. and Hayashizaki,Y.
		High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
		Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES		Location/Qualifiers
source		1..226
		/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="CS7BL/6J"
		/db_xref="taxon:10090"
		/clone="4632413D01"
		/sex="mixed"
		/tissue_type="skin"
		/dev_stages="0 day neonate"
		/lab_host="DH10B"
		/clone_lib="RIKEN full-length enriched, 0 day neonate skin"
		/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTAAATAATTAATCCCTCCCCCCCC sequence [5' GAGAGAGATTCGAGTAAATAATTAATCCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"
ORIGIN		
Query Match		76.8%; Score 19.2; DB 1; Length 226;
Best Local Similarity		87.5%; Pred. No. 5.3e+02;
Matches		21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY		1 AAAAACTCCTGAGGCTTTCAGT 24
Db		125 AAAAAATTCCTGAGGTTTCATT 148
RESULT 19		
AV232251/c		
LOCUS		
DEFINITION		
AV232251		226 bp mRNA linear EST 15-NOV-2001
ACCESSION		musculus cDNA clone 4632413D01 3' similar to AB003502 Mus musculus mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.
VERSION		AV232251
KEYWORDS		EST.
SOURCE		Mus musculus (house mouse)

[illegible]



Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

source  
 Location/Qualifiers  
 1..251  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="5730524116"  
 /sex="mixed"  
 /dev\_stage="8 days embryo"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, 8 days embryo"  
 /notes="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTTCGAGTTAAATTAATTCCTCCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 251;  
 Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTCTGTAGGCTTCAGT 24  
 ||||| ||||| ||||| |||||  
 Db 129 AAAAAATTCCTGTAGGTTTCATT 106

RESULT 24  
 AW240441 251 bp mRNA linear EST 03-APR-2000  
 LOCUS uc34e09.x1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:2811304 3',  
 DEFINITION mRNA sequence.  
 ACCSSION AW240441  
 VERSION AW240441.1 GI:6574193  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 251)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)

REFERENCE  
 AUTHORS Contact: Robert Strausberg, Ph.D.  
 TITLE Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 JOURNAL Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 COMMENT CDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: polyT not found  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 234.

#### FEATURES

source  
 Location/Qualifiers  
 1..251  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="mix FVB/N, C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:2811304"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP Mam5"  
 /notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 251;  
 Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTCTGTAGGCTTCAGT 24  
 ||||| ||||| ||||| |||||  
 Db 123 AAAAAATTCCTGTAGGTTTCATT 146

#### RESULT 25

AV257859/c  
 LOCUS AV257859 252 bp mRNA linear EST 04-NOV-1999  
 DEFINITION AV257859 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4921540C05 3' similar to AB03502 Mus musculus mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.

ACCESSION AV257859  
 VERSION AV257859.1 GI:6245318  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

#### REFERENCE

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Koijima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al. 1999)  
 Unpublished (1999)

#### TITLE

JOURNAL Contact: Yoshihide Hayashizaki  
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: [genome-res@gsc.riken.jp](mailto:genome-res@gsc.riken.jp), [URL:http://genome.gsc.riken.jp/](http://genome.gsc.riken.jp/)  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and





**DEFINITION** AV237384 RIKEN full-length enriched, 10 day neonate skin Mus musculus cDNA clone 4732420L14 3' similar to AB003502 Mus musculus mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.

**ACCESSION** AV237384

**VERSION** AV237384.1 GI:6189896

**KEYWORDS** EST.

**SOURCE** Mus musculus (house mouse)

**ORGANISM** Mus musculus

**REFERENCE** Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

**AUTHORS** Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomihara, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. RIKEN Mouse ESTs (Konno, H., et al. 1999)

**TITLE** RIKEN Mouse ESTs (Konno, H., et al. 1999)

**JOURNAL** Unpublished (1999)

**COMMENT** Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Teurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

**FEATURES**

**source**

1. .255  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="4732420L14"  
 /sex="mixed"  
 /tissue\_type="skin"  
 /dev\_stage="10 days neonate"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, 10 day neonate skin"  
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of

**ORIGIN**

Query Match 76.8%; Score 19.2; DB 1; Length 255;  
 Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCACT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 133 AAAAAATTCCTGTAGGTTTTCATT 110

**RESULT 28**

**LOCUS** AW090751 268 bp mRNA linear EST 15-OCT-1999

**DEFINITION** xc87a06.x1 NCI\_CGAP\_Brn35 Homo sapiens cDNA clone IMAGE:2591218 3', mRNA sequence.

**ACCESSION** AW090751

**VERSION** AW090751.1 GI:6048095

**KEYWORDS** EST.

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 268)

**AUTHORS** NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

**TITLE** National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index

**JOURNAL** Unpublished (1998)

**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@emil.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 254.

**FEATURES**

**source**

1. .268  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2591218"  
 /tissue\_type="tumor, 5 pooled (see description)"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Brn35"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI. Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.33 Kb. Tumor types include: meningioma, oligodendroglioma, astrocytoma (grade II), medulloblastoma, astrocytoma (grade IV). Life Technologies catalog #: 11544-012"

**ORIGIN**

Query Match 76.8%; Score 19.2; DB 1; Length 268;  
 Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCACT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 130 AAAAAATTCCTGTAGGTTTTCATT 153

**RESULT 29**

sequence [5' GAGAGAGAGATTCGAGTTTAATAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"



```

CN471655/c
LOCUS          268 bp      mRNA      linear      EST 22-APR-2004
DEFINITION     hh_Ab_Brain2000_000001758 A. burtoni Cichlid Brain cDNA library
                2000, Russell Fernald Astatotilapia burtoni cDNA clone
                hh_Ab_Brain2000_000001758 5', mRNA sequence.
CN471655
ACCESSION      CN471655.1 GI:46492099
VERSION        EST
KEYWORDS       Astatotilapia burtoni
SOURCE         Astatotilapia burtoni
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
                Labroidae; Cichlidae; African cichlids; Pseudocrenilabrinae;
                Haplochromini; Astatotilapia.
REFERENCE      1 (bases 1 to 268)
AUTHORS       Renn,S.C., Aubin-Horth,N. and Hofmann,H.A.
TITLE         Biologically meaningfull expression profiling across species using
                heterologous hybridization to a cDNA microarray
JOURNAL        BMC Genomics 5 (1), 42 (2004)
PUBMED        15238158
COMMENT       Contact: Hans A. Hofmann
                Bauer Center for Genomics Research
                Harvard University
                7 Divinity Avenue, Cambridge, MA 02138, USA
                Tel: 617-384-8058
                Fax: 617-495-2196
                Email: hans@cgr.harvard.edu
                PCR primers
                FORWARD: CSVP3: AAGCGCGCAATTAAACCTCACTA
                BACKWARD: CSVP2: TTCCGACGTCAGCGTTGTAAAA
                Plate: 019 row: C column: 07
                Seq primer: CSVP3: AAGCGCGCAATTAAACCTCACTA.
FEATURES       Location/Qualifiers
                1..268
                /organism="Astatotilapia burtoni"
                /mol_type="mRNA"
                /db_xref="taxon:8153"
                /clone="hh_Ab_Brain2000_000001758"
                /sex="mixed"
                /tissue_type="brain"
                /dev_stage="mixed"
                /lab_host="X11-Blue MRF, E.coli"
                /clone_lib="A. burtoni Cichlid Brain cDNA library 2000,
                Russell Fernald"
                /note="Vector: pBSISK-; Site 1: EcoRI; Site 2: XhoI; mRNA
                was isolated from the brains of A. burtoni for both sexes
                at all stages of development and reproductive status."
ORIGIN
Query Match          76.8%; Score 19.2; DB 7; Length 268;
Best Local Similarity 87.5%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||
Db 222 AAAAAAGTCTGTGGGCTTTCAT 199

RESULT 30
BB419091/c
LOCUS          269 bp      mRNA      linear      EST 16-JUL-2000
DEFINITION     BB419091 RIKEN full-length enriched, 12 days embryo spinal cord Mus
                musculus cDNA clone C530014L21 3', similar to AB003502 Mus musculus
                mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.
BB419091
ACCESSION      BB419091.1 GI:9240446
VERSION        EST.
KEYWORDS       Mus musculus (house mouse)
SOURCE         Mus musculus
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

---

```

REFERENCE      1 (bases 1 to 269)
AUTHORS       Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
                Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
                Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
                Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
                Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
                Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
                Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
                Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
                Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
                Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
                Watanabe,S., Yamamura,T., Yamanaka,I., Yano,K., Yasunishi,A.,
                Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
                Hayashizaki,Y.
TITLE         RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL        Unpublished (2000)
COMMENT       Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsr.riken.jp, URL:http://genome.gsc.riken.jp/
                Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S.,
                Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Thermostabilization and thermoactivation of thermolabile enzymes by
                trehalose and its application for the synthesis of full length
                cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
                Itoh,M., Kitesunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
                Okazaki,Y. and Hayashizaki,Y.
                Automated filtration-based high-throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
                19-44 (1999)
                Please visit our web site (http://genome.rtc.riken.go.jp) for
                further details.
FEATURES       Location/Qualifiers
                1..269
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="C530014L21"
                /tissue_type="spinal cord"
                /dev_stage="12 days embryo"
                /lab_host="DH10B"
                /clone_lib="RIKEN full-length enriched, 12 days embryo
                spinal_cord"
                /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                prepared and sequenced in Mouse Genome Encyclopedia
                Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in
                RIKEN. Division of Experimental Animal Research in Riken
                contributed to prepare mouse tissues. 1st strand cDNA was
                primed with a primer [5',
                GAGAGAGAGAGATCCAGAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was
                prepared by using trehalose thermo-activated reverse
                transcriptase and subsequently enriched for full-length by
                cap-trapper. Second strand cDNA was prepared with the
                primer adapter of sequence [5',
                GAGAGAGATTCGAGTTAAATTAATTAATCCCCCCCCCC 3']. cDNA
                was cleaved with XhoI and BamHI. Vector: a modified
                pBluescript KS(+) after bulk excision from Lambda FLC I."
ORIGIN
Query Match          76.8%; Score 19.2; DB 2; Length 269;
Best Local Similarity 87.5%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
    ||||| ||||| ||||| ||||| |||||

```

```

Db      147 AAAAAATTCCTAGTGGTTTCATT 124
RESULT 31
LOCUS   AW578173
DEFINITION CM2-CN0039-110100-069-f10 CN0039 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW578173
VERSION   AW578173.1 GI:7253222
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homiidae; Homo.
REFERENCE 1 (bases 1 to 272)
          HCGP http://www.ludwig.org.br/ORESTES.
AUTHORS   The FAPESP/LICR Human Cancer Genome Project
TITLE     Unpublished (1999)
JOURNAL
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-CN0039-
          110100-069-f10&t3=2000-01-11&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 272.
FEATURES             source
     source          1..272
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="CN0039"
                     /note="Organ: colon_normal; Vector: puc18; Site 1: SmaI;
                     Site 2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196, 716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN
Query Match      76.8%; Score 19.2; DB 1; Length 272;
Best Local Similarity 87.5%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy      1 AAAAAATTCCTAGTGGTTTCAGT 24
        ||||| ||||| ||||| |||||
Db      44 AAAAAATTCCTAGTGGTTTCATT 67

RESULT 32
LOCUS   BB367267/c
DEFINITION BB367267 RIKEN full-length enriched, 16 days embryo head Mus
          musculus cDNA clone C130037D05 3', similar to AB003502 Mus musculus
          mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.
ACCESSION BB367267
VERSION   BB367267.1 GI:9079096
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 273)

```

**AUTHORS**

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Wachihi, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE**

RIKEN Mouse ESTs (Konno, H., et al.)

**JOURNAL**

Unpublished (2000)

**COMMENT**

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

**FEATURES**

Location/Qualifiers

1..273  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="C130037D05"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="16 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 16 days embryo head"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCCTAGTGGTTTCAGT 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

**ORIGIN**

Query Match 76.8%; Score 19.2; DB 2; Length 273;  
Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 AAAAACTCTAGGCTTTCAGT 24
DB 150 AAAAAATTCCTGTAGGTTTTCATT 127

RESULT 33
AV028268/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AV028268 274 bp mRNA linear EST 31-AUG-1999
AV028268 Mus musculus adult C57BL/6J liver Mus musculus cDNA clone
1300019H11, mRNA sequence.
AV028268
AV028268.1 GI:4807878
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 274)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Suganara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y.,
Tominaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source 1
Location/Qualifiers
1..274
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1300019H11"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="Mus musculus adult C57BL/6J liver"

ORIGIN
Query Match 76.8%; Score 19.2; DB 1; Length 274;
Best Local Similarity 87.5%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTAGGCTTTCAGT 24
DB 127 AAAAAATTCCTGTAGGTTTTCATT 150

RESULT 35
AV227953/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AV227953 281 bp mRNA linear EST 14-NOV-2001
AV227953 RIKEN full-length enriched, 14 days embryo liver Mus
musculus cDNA clone 443241J23 3' similar to AB003502 Mus musculus
mRNA for Guanine Nucleotide Regulatory Protein, mRNA sequence.
AV227953
AV227953.1 GI:6180472
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 281)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,
Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tominaga, N.,
Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki

```





cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Cloned through the NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert length: 1060 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 286.  
Location/Qualifiers  
1. .292  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2499843"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Brn25"  
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTGAAGTGGAGCGCCGATAGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 292;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||||| |||||  
DB 125 AAAAAATTCCTGTAGGTTTTCATT 148

RESULT 40  
AI801161  
LOCUS  
DEFINITION  
293 bp mRNA linear EST 06-JUL-1999  
to79e06.x1 NCI\_CGAP\_Gas4 Homo sapiens cDNA clone IMAGE:2184514 3',  
mRNA sequence.  
AI801161  
VERSION  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Cloned through the NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 265.  
Location/Qualifiers  
1. .293  
/organism="Homo sapiens"  
/mol\_type="mRNA"

## FEATURES

source  
1. .293  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/clone="IMAGE:2184514"  
/tissue\_type="poorly differentiated adenocarcinoma with signet ring cell features"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Gas4"  
/note="Organ: stomach; Vector: pCMV-SPORT6; Site.1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 293;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||||| |||||  
DB 125 AAAAAATTCCTGTAGGTTTTCATT 148

## RESULT 41

AW273587  
LOCUS  
DEFINITION  
294 bp mRNA linear EST 03-JAN-2000  
xv67f07.x1 NCI\_CGAP\_Lu28 Homo sapiens cDNA clone IMAGE:2818213 3',  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

## REFERENCE

AUTHORS  
TITLE  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Cloned through the NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 293.  
Location/Qualifiers  
1. .294  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2818213"  
/tissue\_type="two pooled squamous cell carcinomas"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Lu28"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site.1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

## FEATURES

source  
1. .294  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2818213"  
/tissue\_type="two pooled squamous cell carcinomas"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Lu28"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site.1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 294;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||||| |||||  
DB 123 AAAAAATTCCTGTAGGTTTTCATT 146







to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao (NIA)."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 298;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGT 24  
||||| ||||||| ||||||| |||||||  
Db 123 AAAAAAATCTCTAGGCTTTTCATT 146

## RESULT 44

AA450256

## LOCUS

DEFINITION z42g02.s1 Soares total\_fetus\_Nb2HF8\_9w Homo sapiens cDNA clone  
IMAGS:789170 3', mRNA sequence.

## ACCESSION

AA450256

## VERSION

AA450256.1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 300)

Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,

Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J.,

Moore B., Schellenberg K., Steptoe M., Tan F., Theising B.,

White Y., Wyllie T., Waterston R. and Wilson R.

WashU-Merck EST Project 1997

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 291.

Location/Qualifiers

1. .300

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:5986787"

/db\_xref="taxon:9606"

/clone="IMAGE:789170"

/dev\_stage="8-9 weeks"

/lab\_host="DH10B"

/clone\_lib="Soares total\_fetus\_Nb2HF8\_9w"

/note="Vector: p7T73D-Fac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAACTGAGTGGGCGCGCTTAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified p7T73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo. "

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 300;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCTCTAGGCTTTTCAGT 24  
||||| ||||||| ||||||| |||||||  
Db 123 AAAAAAATCTCTAGGCTTTTCATT 146

## RESULT 45

BM248034

## LOCUS

DEFINITION BM248034 KIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus  
musculus cDNA clone NIA:K0810F03 IMAGE:30079358 3', mRNA sequence.

## ACCESSION

BM248034

## VERSION

BM248034.2

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 300)

Piao Y., Ko N.T., Lim M.K. and Ko M.S.H.

Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

Genome Res. 11 (9), 1553-1558 (2001)

11544199

On Dec 17, 2001 this sequence version replaced gi:17883304.

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Caselli Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@leuun.grc.nia.nih.gov

Seq primer: -21M13 Forward

High quality sequence stop: 300

POLYA=yes.

Location/Qualifiers

1. .300

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="niaEST:K0810F03-3"

/db\_xref="taxon:10090"

/clone="NIA:K0810F03 IMAGE:30079358"

/tissue\_type="whole embryo including extraembryonic

tissues at 8.5-days postcoitum"

/dev\_stage="8.5-days postcoitum"

/lab\_host="DH10B"

/clone\_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library

(Long)"

/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (http://leuun.grc.nia.nih.gov/cDNA). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11544199]]. Total RNAs were

extracted from a pool of 13 embryos at 8.5-days

postcoitum. Double-stranded cDNAs were synthesized with an

Oligo(dT) primer [Invitrogen].

5'-PGACTAGTCTAGATCGAGCGCGCTTTT-3'] from

9.1 ug of total RNA, treated with T4 DNA polymerase, and

purified by ethanol-precipitation. The cDNAs were ligated

to Lone-linker LL-Sal4, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Taq polymerase (Takara) with a primer Sal4-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pSPORT1 plasmid vector.



Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

# TITLE

## JOURNAL COMMENT

The WashU-HMI Mouse EST Project  
 Contact: Marra M/Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:324230

Putative full length read

vector to vector length is 390

Seq primer: -40m13 ET

High quality sequence stop: 310.

## FEATURES

source  
 1..311  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6 x DBA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:537294"  
 /sex="pooled"  
 /tissue\_type="embryo"  
 /dev\_stage="7.5dpc"  
 /lab\_host="DH12S"  
 /clone\_lib="Beddington mouse embryonic region"  
 /note="Organ: whole embryo; Vector: pCMV-SPORT; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Gastrulating embryos were collected at 7.5dpc from C57BL/6 x DBA matings, excluding embryos that had developed head folds and all extraembryonic tissues. Average insert size: 1.3 kb (range: 0.5 - 3.0 kb). Referenced in Development 121, 2479-2489 (1995)"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 311;  
 Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 AAAAACTCTGTAGGCTTTTCAGT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 267 AAAAAATTCCTGTAGGTTTCATT 244

## RESULT 49

AW323152 312 bp mRNA linear EST 26-JAN-2000  
 LOCUS uc58c11.x1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:2646740 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AW323152  
 VERSION AW323152.1 GI:6757177  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 312)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

## FEATURES

source

MGI:1027192  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 311.  
 Location/Qualifiers  
 1..312  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:2646740"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 Stem cell origin.  
 /lab\_host="DH108"  
 /clone\_lib="NCI\_CGAP Lu29"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 1; Length 312;  
 Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 AAAAACTCTGTAGGCTTTTCAGT 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 126 AAAAAATTCCTGTAGGTTTCATT 149

## RESULT 50

CX846858 313 bp mRNA linear EST 02-FEB-2005  
 LOCUS JGI\_CAAK12911.rev NIH XGC tropBrn3 Xenopus tropicalis cDNA clone  
 DEFINITION IMAGE:7793462 3', mRNA sequence.  
 ACCESSION CX846858  
 VERSION CX846858.1 GI:58502125  
 KEYWORDS EST.  
 SOURCE Xenopus tropicalis (western clawed frog)  
 ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.

## REFERENCE

AUTHORS

Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., NG, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

Other ESTs: JGI\_CAAK12911.fwd

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,

University of California, Berkeley:

http://tropicalis.berkeley.edu/home)

cDNA Library Preparation: DOE Joint Genome Institute:

http://www.jgi.doe.gov

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix 'rev'

indicates a reverse sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Poly-A: Based upon the presence of a run of 14 or more T residues

at the beginning of the sequence, this clone was polyadenylated.

The resulting Poly-T sequence has been removed.

Plate: CAAK 0133 row: m column: 12  
High quality sequence stop: 311  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1..313  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:7793462"  
/tissue\_type="Brain"  
/dev\_stage="Adult"  
/lab\_host="ElectroMAX DH10B"  
/clone\_lib="NIH\_XGC\_tropBrn3"  
/note="Vector: pCMVSPORT6; Site\_1: SalI; Site\_2: NotI;  
This library was made from dr primed cDNA and cloned into  
Invitrogen pCMVSPORT6 vector. The work was done at DOE  
Joint Genome Institute. Poly A RNA were primed with 5',  
GACTAGTCTAGATCGCAG CGGCGCCCTTTT TTTT 3'. cDNA  
were ligated to SalI adapter (5' TCGACCCACGCGTCGG and  
5'CGACGCGTGGG), digested with NotI, size fractionated in  
1.1% agarose gel electrophoresis and ligated into NotI and  
SalI digested pCMVSPORT6 vector."

ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 313;  
Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGCTCTGTAGGCTTTTCAGT 24  
||||| ||||||||| |||||  
Db 252 AAAACATCTGTAGGCTTCAGT 275

Search completed: February 3, 2006, 22:02:21  
Job time : 2960.67 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-20

Perfect score: 25

Sequence: 1 aaaaactctgtaggcttcagt 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6S COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfilese1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	76.8	2584	3	US-09-949-016-5287
C 2	19.2	76.8	47284	3	US-09-949-016-17029
C 3	18.8	75.2	23776	3	US-09-949-002-702
4	18.6	74.4	601	3	US-09-949-016-76001
5	18.6	74.4	601	3	US-09-949-016-76002
6	18.6	74.4	39982	3	US-09-820-924-3
7	18.6	74.4	39982	3	US-10-369-626-3
8	18.6	74.4	51043	3	US-09-949-016-12739
9	18.6	74.4	51046	3	US-09-949-016-13946
10	17.6	70.4	1563	2	US-08-279-590A-1
11	17.6	70.4	1563	3	US-08-910-092-1
12	17.6	70.4	1657	3	US-09-270-767-9840
13	17.6	70.4	1867	3	US-08-630-915A-13
14	17.6	70.4	1867	3	US-09-879-357-13
15	17.6	70.4	8511	3	US-09-949-016-14976
C 16	17.6	70.4	36816	3	US-09-949-016-16218
C 17	17.6	70.4	118143	3	US-09-949-016-17196
C 18	17.2	68.8	601	3	US-09-949-016-144942
19	17.2	68.8	601	3	US-09-949-016-144943
C 20	17.2	68.8	1971	3	US-09-620-312D-960
C 21	17.2	68.8	3914	3	US-10-104-047-850
C 22	17.2	68.8	52667	3	US-09-949-016-12019
C 23	17.2	68.8	146307	3	US-09-949-016-14881
C 24	17.2	68.8	146307	3	US-09-949-016-14882

25	17.2	68.8	146307	3	US-09-949-016-14883	Sequence 14883, A
26	17.2	68.8	146307	3	US-09-949-016-14884	Sequence 14884, A
27	17.2	68.8	146307	3	US-09-949-016-14885	Sequence 14885, A
28	17.2	68.8	146307	3	US-09-949-016-14886	Sequence 14886, A
29	17.2	68.8	146307	3	US-09-949-016-14887	Sequence 14887, A
30	17.2	68.8	146307	3	US-09-949-016-14888	Sequence 14888, A
31	17.2	68.8	148405	3	US-09-949-016-11747	Sequence 11747, A
32	17.2	68.8	148405	3	US-09-949-016-12835	Sequence 12835, A
33	17.2	68.8	148405	3	US-09-949-016-12836	Sequence 12836, A
34	17.2	68.8	148405	3	US-09-949-016-12837	Sequence 12837, A
C 35	17.2	68.8	169998	3	US-09-676-610B-24	Sequence 24, Appl
C 36	17.2	68.8	177251	3	US-09-949-016-15841	Sequence 15841, A
C 37	17.2	68.8	177496	3	US-09-877-177A-10	Sequence 10, Appl
C 38	17	68.0	601	3	US-09-949-016-133546	Sequence 133546, A
C 39	17	68.0	601	3	US-09-949-016-133547	Sequence 133547, A
40	17	68.0	703	3	US-09-270-767-982	Sequence 982, App
41	17	68.0	703	3	US-09-270-767-16264	Sequence 16264, A
C 42	17	68.0	1095	3	US-09-134-000C-625	Sequence 625, App
C 43	17	68.0	1939	3	US-09-799-451-566	Sequence 566, App
C 44	17	68.0	2205	3	US-09-949-016-1046	Sequence 1046, App
C 45	17	68.0	3416	3	US-09-949-016-4827	Sequence 4827, App
C 46	17	68.0	3416	3	US-09-949-016-4828	Sequence 4828, App
C 47	17	68.0	3490	3	US-09-949-016-4829	Sequence 4829, App
C 48	17	68.0	3490	3	US-09-949-016-4830	Sequence 4830, App
49	17	68.0	17132	3	US-09-949-016-15361	Sequence 15361, A
50	17	68.0	24150	3	US-09-949-016-12438	Sequence 12438, A
C 51	17	68.0	30271	3	US-09-949-016-12796	Sequence 12796, A
C 52	17	68.0	30272	3	US-09-949-016-14006	Sequence 14006, A
C 53	17	68.0	36075	3	US-09-949-016-16571	Sequence 16571, A
C 54	17	68.0	36075	3	US-09-949-016-16572	Sequence 16572, A
C 55	17	68.0	36625	3	US-09-949-016-12788	Sequence 12788, A
C 56	17	68.0	37133	3	US-09-949-016-16569	Sequence 16569, A
C 57	17	68.0	37133	3	US-09-949-016-16570	Sequence 16570, A
C 58	17	68.0	202111	3	US-09-949-016-13877	Sequence 13877, A
C 59	17	68.0	245286	3	US-09-949-016-15497	Sequence 15497, A
60	17	68.0	256171	3	US-09-949-016-12822	Sequence 12822, A
61	17	68.0	256176	3	US-09-949-016-12824	Sequence 12824, A
62	17	68.0	451924	3	US-09-949-016-12896	Sequence 12896, A
63	17	68.0	451925	3	US-09-949-016-17305	Sequence 17305, A
64	16.8	67.2	442	3	US-09-270-767-12773	Sequence 12773, A
C 65	16.8	67.2	1589	3	US-09-799-451-637	Sequence 637, App
C 66	16.8	67.2	8341	2	US-08-737-825-1	Sequence 1, Appli
C 67	16.6	66.4	306	3	US-09-513-999C-8250	Sequence 8250, App
68	16.6	66.4	316	3	US-09-662-402A-28	Sequence 28, Appl
69	16.6	66.4	601	3	US-09-949-016-31372	Sequence 31372, A
70	16.6	66.4	601	3	US-09-949-016-76138	Sequence 76138, A
71	16.6	66.4	601	3	US-09-949-016-12235	Sequence 12235, A
72	16.6	66.4	601	3	US-09-949-016-135598	Sequence 135598, A
C 73	16.6	66.4	601	3	US-09-949-016-148083	Sequence 148083, A
C 74	16.6	66.4	601	3	US-09-949-016-148189	Sequence 148189, A
C 75	16.6	66.4	601	3	US-09-949-016-204052	Sequence 204052, A
C 76	16.6	66.4	2660	3	US-09-949-016-2206	Sequence 2206, App
77	16.6	66.4	2669	3	US-09-949-016-691	Sequence 691, App
78	16.6	66.4	2674	2	US-08-926-724-2	Sequence 2, Appli
C 79	16.6	66.4	4756	3	US-09-221-017B-913	Sequence 913, App
C 80	16.6	66.4	4530	3	US-09-949-016-15185	Sequence 15185, A
C 81	16.6	66.4	19299	3	US-09-949-016-13014	Sequence 13014, A
C 82	16.6	66.4	52523	3	US-09-949-016-12433	Sequence 12433, A
C 83	16.6	66.4	52530	3	US-09-949-016-12433	Sequence 12433, A
C 84	16.6	66.4	98701	3	US-09-949-016-15988	Sequence 15988, A
C 85	16.6	66.4	98701	3	US-09-949-016-15989	Sequence 15989, A
C 86	16.6	66.4	129380	3	US-09-949-016-12544	Sequence 12544, A
C 87	16.6	66.4	139257	3	US-09-920-671-11	Sequence 11, Appl
C 88	16.6	66.4	161124	3	US-09-949-016-11760	Sequence 11760, A
C 89	16.6	66.4	187595	3	US-09-949-016-1546	Sequence 1546, A
90	16.6	66.4	276237	3	US-09-949-016-17504	Sequence 17504, A
91	16.6	66.4	392000	3	US-10-027-983-11	Sequence 11, Appl
C 92	16.4	65.6	269	3	US-09-513-999C-9119	Sequence 9119, App
C 93	16.4	65.6	601	3	US-09-949-016-149843	Sequence 149843, A
94	16.4	65.6	89450	3	US-09-949-016-15848	Sequence 15848, A
C 95	16.4	65.6	105050	3	US-09-949-016-15953	Sequence 15953, A
C 96	16.2	64.8	450	3	US-09-513-999C-28047	Sequence 28047, A
C 97	16.2	64.8	601	3	US-09-949-016-32668	Sequence 32668, A



US-09-949-002-702

Query Match 75.2%; Score 18.8; DB 3; Length 23776;  
Best Local Similarity 90.9%; Pred. No. 33;  
Matches 20: Conservative 0; Mismatches 2; Indels 0;

**Qy**

4 AAACCTCCTGTAGGCTTTCAGTG 25  
||| ||| ||| ||| ||| ||| ||| |||  
**Db**

1282 AAAATCCTGGAGGCTTTCAGTG 1261

## RESULT 4

```

US-09-949-016-76001
; Sequence 76001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION
; AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76001
; LENGTH: 601

```

; LENGTH: 601  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-76001

Query Match 74.4%; Score 18.6; DB 3; Length 601;  
Best Local Similarity 84.0%; Pred. No. 16;  
Matches 21; Conservative 0; Mismatches 4; Indels

QY 1 AAAAACTCCTGTAGGCTTTCAGTG 25  
||| ||| ||| ||| ||| ||| ||| |||  
Db 442 AAAGAAATCCTGTTGGCTTTCGTG 466

RESIST

```

US-09-949-016-76002
; Sequence 76002, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76002
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-76002

```

US-09-949-016-76002

Query Match 74.4%; Score 18.6; DB 3; Length 601;  
Best Local Similarity 84.0%; Pred. No. 16;  
Matches 21: Conservative 0; Mismatches 4; Indels

**Qy**

1 AAAAAACTCCTGTAGGCTTTCAGTG 25  
|||||

**Dβ**

61 AAAGAAATCCTGTTGGCTTTCGTG 85  
|||||

## RESULT 6

```

US-09-820-924-3
; Sequence 3, Application US/09820924
; Patent No. 6553151
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCES: CL001213
; CURRENT APPLICATION NUMBER: US/09/820,924
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 39982
; TYPE: DNA
; ORGANISM: Human
US-09-820-924-3

```

Query Match	74.4%;	Score 18.6;	DB 3;	Length 39982;
Best Local Similarity	84.0%;	Pred. No. 48;		
Matches	21;	Conservative	0;	Mismatches 4;
				Indels 0;

Qy 1 AAAAAACTCTGTAGGCTTTCAGTG 25  
||| ||| ||| ||| ||| ||| ||| |||  
Db 14081 AAAGAAATCCCTGTGGCTTCTGTG 14

## RESULT 7

```

US-10-369-626-3
; Sequence 3, Application US/10369626
; Patent No. 6787344
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001213DIV
; CURRENT APPLICATION NUMBER: US/10/369,626
; CURRENT FILING DATE: 2003-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: fastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 39982
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-369-626-3

```

Query Match	74.4%;	Score 18.6;	DB 3;	Length 39982;
Best Local Similarity	84.0%;	Pred. No. 48;		
Matches 21.	Conservative	0: Mismatches	4: Indels	0

Qy 1 AAAAACTCCTGTAGGCTTTCAGTG 25  
|||  
Db 14081 AAGCAATCCTGTGGCTTCTGT 14

## RESULT 8

RECD. 949-016-12739  
 US-09-949-016-12739  
 ; Sequence 12739, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307



; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12739  
; LENGTH: 51043  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(51043)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12739

Query Match 74.4%; Score 18.6; DB 3; Length 51043;  
Best Local Similarity 84.0%; Pred. No. 51;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGTG 25

Db 13899 AAAGAAATCCTGTGGCTTTTCGTG 13923

## RESULT 9

US-09-949-016-13946  
; Sequence 13946, Application US/09949016  
; Patent No. 6812339

## GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13946  
; LENGTH: 51046  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(51046)  
; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-13946

Query Match 74.4%; Score 18.6; DB 3; Length 51046;  
Best Local Similarity 84.0%; Pred. No. 51;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGTG 25

Db 13899 AAAGAAATCCTGTGGCTTTTCGTG 13923

## RESULT 10

US-08-279-590A-1  
; Sequence 1, Application US/08279590A  
; Patent No. 5656749  
; GENERAL INFORMATION:  
; APPLICANT: ELIOT R. SPINDEL, SRINIVASA

; APPLICANT: NAGALLA AND BRENDA BARRY  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING  
; RECEPTORS FOR BOMBESIN-LIKE  
; PEPTIDES  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/279,590A  
; FILING DATE: July 22, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: No. 5656749e  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Y. Rocky Tsao  
; REGISTRATION NUMBER: 34,053  
; REFERENCE/DOCKET NUMBER: 00537/098001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1563  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-279-590A-1

Query Match 70.4%; Score 17.6; DB 2; Length 1563;  
Best Local Similarity 83.3%; Pred. No. 66;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24

Db 187 AAAAAATCGCTGAGGTTTCAGT 210

## RESULT 11

US-08-910-092-1  
; Sequence 1, Application US/08910092  
; Patent No. 5814463  
; GENERAL INFORMATION:  
; APPLICANT: Spindel, Eliot R.  
; APPLICANT: Nagalla, Srinivasa  
; APPLICANT: Barry, Brenda

; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING RECEPTORS  
; FOR BOMBESIN-LIKE PEPTIDES  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910,092  
FILING DATE: 12-AUG-1997  
PRIOR APPLICATION DATA: 08/279,590  
FILING DATE: 22-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsao, Y. Rocky  
REGISTRATION NUMBER: 34,053  
REFERENCE/DOCKET NUMBER: 00537/098002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1563 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 192...1319  
US-08-910-092-1

Query Match 70.4%; Score 17.6; DB 2; Length 1563;  
Best Local Similarity 83.3%; Pred. No. 66;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 187 AAAAAATGCTGAAGGTTTTCAGT 210

## RESULT 12

US-09-270-767-9840  
Sequence 9840, Application US/09270767  
Patent No. 6703491  
GENERAL INFORMATION:  
APPLICANT: Homburger et al.  
TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9840  
LENGTH: 1657  
TYPE: DNA  
ORGANISM: *Drosophila melanogaster*  
US-09-270-767-9840

Query Match 70.4%; Score 17.6; DB 3; Length 1657;  
Best Local Similarity 83.3%; Pred. No. 67;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 550 AGAAAACTCTGTGGGCTATCAGT 573

## RESULT 13

US-08-630-915A-13  
Sequence 13, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLES, Dana M.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME

NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-630-915A-13

Query Match 70.4%; Score 17.6; DB 3; Length 1867;  
Best Local Similarity 83.3%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 750 ATAAAACTCTGTGGCTTTTCAGT 773

## RESULT 14

US-09-879-957-13  
Sequence 13, Application US/09879957  
Patent No. 6709821  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6709821h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME

NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-879-957-13

Query Match 70.4%; Score 17.6; DB 3; Length 1867;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
DB 750 ATAAAGTCTGTGGCTTTGAGT 773

RESULT 15
US-09-949-016-14976
; Sequence 14976, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14976
; LENGTH: 8511
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14976

Query Match 70.4%; Score 17.6; DB 3; Length 8511;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
DB 5581 AAAAACTCTGTATATCTTTTCATT 5604

RESULT 16
US-09-949-016-16218/c
; Sequence 16218, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16218
; LENGTH: 36816
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16218

Query Match 70.4%; Score 17.6; DB 3; Length 36816;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
DB 36272 AACATACTCTGTAGTCTTTTCATT 36249

RESULT 17
US-09-949-016-17196/c
; Sequence 17196, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17196
; LENGTH: 118143
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17196

Query Match 70.4%; Score 17.6; DB 3; Length 118143;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
DB 8436 AAAAAATCCTATAGCATTCAGT 8413

RESULT 18
US-09-949-016-144942
; Sequence 144942, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
```

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 144942  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-144942

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 82;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
|||||  
DB 369 AAAAACTCTGTAAACATTCA 390

## RESULT 19

US-09-949-016-144943  
; Sequence 144943, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 144943  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-144943

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 82;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
|||||  
DB 2 AAAAACTCTGTAAACATTCA 23

## RESULT 20

US-09-620-312D-960/c  
; Sequence 960, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 960  
; LENGTH: 1971  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1884)  
US-09-620-312D-960

Query Match 68.8%; Score 17.2; DB 3; Length 1971;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
|||||  
DB 1403 AAAAAACCTGTAAAGCTTTGA 1382

## RESULT 21

US-10-104-047-850/c  
; Sequence 850, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 850  
; LENGTH: 3914  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-850

Query Match 68.8%; Score 17.2; DB 3; Length 3914;  
Best Local Similarity 86.4%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
|||||  
DB 1031 AAAAACTCTGTGGCCTTTCA 1010

## RESULT 22

US-09-949-016-12019/c  
; Sequence 12019, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755

RESULT 24  
US-09-949-016-14882  
; Sequence 14882, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

```

RESULT 26
US-09-949-016-14884
; Sequence 14884, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14884
; LENGTH: 146307
; TYPE: DNA

```

```

; ORGANISM: Human
US-09-949-016-14884

Query Match          68.8%; Score 17.2; DB 3; Length 146307;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 27
US-09-949-016-14885
; Sequence 14885, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14885
; LENGTH: 146307
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14885

Query Match          68.8%; Score 17.2; DB 3; Length 146307;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 28
US-09-949-016-14886
; Sequence 14886, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14886
; LENGTH: 146307
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14886

Query Match          68.8%; Score 17.2; DB 3; Length 146307;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 29
US-09-949-016-14887
; Sequence 14887, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14887
; LENGTH: 146307
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14887

Query Match          68.8%; Score 17.2; DB 3; Length 146307;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 30
US-09-949-016-14888
; Sequence 14888, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14888
; LENGTH: 146307
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14888

Query Match          68.8%; Score 17.2; DB 3; Length 146307;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452
```

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12836
; LENGTH: 148405
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12836

Query Match      68.8%; Score 17.2; DB 3; Length 148405;
Best Local Similarity 86.4%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 34
US-09-949-016-12837
; Sequence 12837, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12837
; LENGTH: 148405
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12837

Query Match      68.8%; Score 17.2; DB 3; Length 148405;
Best Local Similarity 86.4%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCTCTGTAGGCTTTTCAG 23
Db 18431 AAGACCTCTGTGGCTTTTCAG 18452

RESULT 35
US-09-676-610B-24/C
; Sequence 24, Application US/09676610B
; Patent No. 644465
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Freier

```



;; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION  
;; FILE REFERENCE: RTS-0138  
;; CURRENT APPLICATION NUMBER: US/09/676,610B  
;; CURRENT FILING DATE: 2000-09-29  
;; NUMBER OF SEQ ID NOS: 182  
;; SEQ ID NO 24  
;; LENGTH: 16998  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: exon  
;; LOCATION: (1208)...(1472)  
;; NAME/KEY: intron  
;; LOCATION: (1473)...(124390)  
;; NAME/KEY: exon  
;; LOCATION: (124391)...(124544)  
;; NAME/KEY: intron  
;; LOCATION: (124545)...(125409)  
;; NAME/KEY: exon  
;; LOCATION: (125410)...(125595)  
;; NAME/KEY: intron  
;; LOCATION: (125596)...(128711)  
;; NAME/KEY: exon  
;; LOCATION: (128712)...(128848)  
;; NAME/KEY: intron  
;; LOCATION: (128849)...(133400)  
;; NAME/KEY: exon  
;; LOCATION: (133401)...(133469)  
;; NAME/KEY: intron  
;; LOCATION: (133470)...(134652)  
;; NAME/KEY: exon  
;; LOCATION: (134653)...(134773)  
;; NAME/KEY: intron  
;; LOCATION: (134774)...(136116)  
;; NAME/KEY: exon  
;; LOCATION: (136117)...(136261)  
;; NAME/KEY: intron  
;; LOCATION: (136262)...(137936)  
;; NAME/KEY: exon  
;; LOCATION: (137937)...(138053)  
;; NAME/KEY: intron  
;; LOCATION: (138054)...(138637)  
;; NAME/KEY: exon  
;; LOCATION: (138638)...(138766)  
;; NAME/KEY: intron  
;; LOCATION: (138767)...(138864)  
;; NAME/KEY: exon  
;; LOCATION: (138865)...(138940)  
;; NAME/KEY: intron  
;; LOCATION: (138941)...(139765)  
;; NAME/KEY: exon  
;; LOCATION: (139766)...(139860)  
;; NAME/KEY: intron  
;; LOCATION: (139861)...(142245)  
;; NAME/KEY: exon  
;; LOCATION: (142246)...(142445)  
;; NAME/KEY: intron  
;; LOCATION: (142446)...(143605)  
;; NAME/KEY: exon  
;; LOCATION: (143606)...(143738)  
;; NAME/KEY: intron  
;; LOCATION: (143739)...(145838)  
;; NAME/KEY: exon  
;; LOCATION: (145839)...(145931)  
;; NAME/KEY: intron  
;; LOCATION: (145932)...(147385)  
;; NAME/KEY: exon  
;; LOCATION: (147386)...(147544)  
;; NAME/KEY: intron  
;; LOCATION: (147545)...(153274)  
;; NAME/KEY: exon  
;; LOCATION: (153275)...(153321)  
;; NAME/KEY: intron

;; LOCATION: (153322)...(155088)  
;; NAME/KEY: exon  
;; LOCATION: (155089)...(155231)  
;; NAME/KEY: intron  
;; LOCATION: (155232)...(156025)  
;; NAME/KEY: exon  
;; LOCATION: (156026)...(156151)  
;; NAME/KEY: intron  
;; LOCATION: (156152)...(156826)  
;; NAME/KEY: exon  
;; LOCATION: (156827)...(156928)  
;; NAME/KEY: intron  
;; LOCATION: (156929)...(163399)  
;; NAME/KEY: exon  
;; LOCATION: (163400)...(163586)  
US-09-676-610B-24

Query Match 68.8%; Score 17.2; DB 3; Length 16998;  
Best Local Similarity 86.4%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCTGTAGGCTTTTCAG 23  
|||||  
Db 10676 AAAATCTCTGAAAGCTTTTCAG 10655

RESULT 36

US-09-949-016-15841  
;; Sequence 15841, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 15841  
;; LENGTH: 177251  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-15841

Query Match 68.8%; Score 17.2; DB 3; Length 177251;  
Best Local Similarity 86.4%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTC 22  
|||||  
Db 79470 AAAAACTCTGTAAACATTC 79491

RESULT 37

US-09-877-177A-10/c  
;; Sequence 10, Application US/09877177A  
;; Patent No. 6582919  
;; GENERAL INFORMATION:  
;; APPLICANT: K. Danenberg  
;; TITLE OF INVENTION: Method of determining Epidermal Growth  
;; Factor Receptor and HER2-Neu Gene Expression  
;; TITLE OF INVENTION: and Correlation of Levels Thereof With Survival  
;; FILE REFERENCE: 11220/120  
;; CURRENT APPLICATION NUMBER: US/09/877,177A  
;; CURRENT FILING DATE: 2001-06-11  
;; NUMBER OF SEQ ID NOS: 11

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-877-177A-10

Query Match          68.8%; Score 17.2; DB 3; Length 197496;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTCCTGTAGGCTTTTCAG 23
    ||||| ||||| ||||| ||||| ||
Db 18676 AAAATCTCTGAAGCTTTTCAG 18655

RESULT 38
US-09-949-016-133546/c
; Sequence 133546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133546
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-133546

Query Match          68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 25
    ||||| ||||| ||||| ||||| ||
Db 428 AAAAAACAATTGTAGCTTTTCATG 404

RESULT 39
US-09-949-016-133547/c
; Sequence 133547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133547
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
```

```
US-09-949-016-133547

Query Match          68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 25
    ||||| ||||| ||||| ||||| ||
Db 406 AAAAAACAATTGTAGCTTTTCATG 382

RESULT 40
US-09-270-767-982
; Sequence 982, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 982
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-982

Query Match          68.0%; Score 17; DB 3; Length 703;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 25
    ||||| ||||| ||||| ||||| ||
Db 560 AAAAAAATGCTTTAAGCTTTTCATTG 584

RESULT 41
US-09-270-767-16264
; Sequence 16264, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16264
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16264

Query Match          68.0%; Score 17; DB 3; Length 703;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 25
    ||||| ||||| ||||| ||||| ||
Db 560 AAAAAAATGCTTTAAGCTTTTCATTG 584

RESULT 42
US-09-134-000C-625/c
; Sequence 625, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
```



```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4828
; LENGTH: 3416
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4828

Query Match      68.0%; Score 17; DB 3; Length 3416;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25
   ||||| ||| || |||||
Db 586 AAAAACTCTGCTGGATTTCAGTG 562

RESULT 47
US-09-949-016-4829/c
; Sequence 4829, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4829
; LENGTH: 3490
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4829

Query Match      68.0%; Score 17; DB 3; Length 3490;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25
   ||||| ||| || |||||
Db 586 AAAAACTCTGCTGGATTTCAGTG 562

RESULT 48
US-09-949-016-4830/c
; Sequence 4830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4830
; LENGTH: 3490
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4830

Query Match      68.0%; Score 17; DB 3; Length 3490;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25
   ||||| ||| || |||||
Db 6872 AAAAAATGCTGATGGCTTTCAGAG 6896

RESULT 50
US-09-949-016-12438
; Sequence 12438, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4830
; LENGTH: 3490
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4830

Query Match      68.0%; Score 17; DB 3; Length 3490;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25
   ||||| ||| || |||||
Db 586 AAAAACTCTGCTGGATTTCAGTG 562

RESULT 49
US-09-949-016-15361
; Sequence 15361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15361
; LENGTH: 17132
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)...(17132)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15361

Query Match      68.0%; Score 17; DB 3; Length 17132;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25
   ||||| ||| || |||||
Db 6872 AAAAAATGCTGATGGCTTTCAGAG 6896

RESULT 50
US-09-949-016-12438
; Sequence 12438, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12438
; LENGTH: 24150
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(24150)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12438

Query Match      68.0%; Score 17; DB 3; Length 24150;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAATCCTGTAGGCTTTCAGTG 25
        ||||| ||||| ||||| |||||
DB      6872 AAAAAATGCTGTAGGCTTTCAGAG 6896

Search completed: February 3, 2006, 16:32:30
Job time : 92.6667 secs
```

**HIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-20

Perfect score: 25

Sequence: 1 aaaaactctgtaggcttcagtg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	8	US-10-719-900-20
C 2	25	100.0	74868	5	US-10-175-523-67
C 3	25	100.0	74868	10	US-11-099-266-67
C 4	23.4	93.6	25	8	US-10-719-900-19
C 5	19.8	79.2	334	8	US-10-425-115-33079
C 6	19.8	79.2	2243	8	US-10-788-792-104
C 7	19.2	76.8	258	7	US-10-242-535A-7089
C 8	19.2	76.8	258	7	US-10-085-783A-7089
C 9	19.2	76.8	289	7	US-10-242-535A-38667
C 10	19.2	76.8	289	7	US-10-085-783A-38667
C 11	19.2	76.8	439	6	US-10-125-968-712
C 12	19.2	76.8	546	5	US-10-027-632-199786
C 13	19.2	76.8	546	6	US-10-027-632-199786
C 14	19.2	76.8	562	3	US-09-864-761-13669
C 15	19.2	76.8	693	4	US-09-925-065A-81129
C 16	19.2	76.8	1400	9	US-10-956-157-6011
C 17	19.2	76.8	2527	6	US-10-108-260A-731
C 18	19.2	76.8	2587	5	US-10-146-473-6
C 19	19.2	76.8	2587	7	US-10-717-597-125
C 20	19.2	76.8	2587	9	US-10-956-157-776
C 21	19.2	76.8	2587	9	US-10-981-277-8
C 22	19.2	76.8	3886	6	US-10-252-157-57
C 23	19.2	76.8	5569	3	US-09-814-353-20129

C	24	19.2	76.8	5570	8	US-10-357-930-24829	Sequence 24829, A
	25	19.2	76.8	150525	9	US-10-981-277-36	Sequence 36, Appl
	26	19.2	76.8	166536	9	US-10-981-277-35	Sequence 35, Appl
	27	18.6	74.4	594	5	US-10-027-632-129877	Sequence 129877, A
	28	18.6	74.4	594	5	US-10-027-632-129877	Sequence 129878, A
	29	18.6	74.4	594	6	US-10-027-632-129877	Sequence 129877, A
	30	18.6	74.4	594	6	US-10-027-632-129877	Sequence 129878, A
	31	18.6	74.4	823	5	US-10-027-632-131274	Sequence 131274, A
	32	18.6	74.4	823	5	US-10-027-632-131275	Sequence 131275, A
	33	18.6	74.4	823	6	US-10-027-632-131274	Sequence 131274, A
	34	18.6	74.4	823	6	US-10-027-632-131275	Sequence 131275, A
	35	18.6	74.4	39982	6	US-10-369-626-3	Sequence 3, Appl
	36	18.6	74.4	39982	8	US-10-900-420-3	Sequence 3, Appl
C	37	18.2	72.8	317	8	US-10-425-115-172275	Sequence 172275, A
C	38	18.2	72.8	351	8	US-10-425-115-33082	Sequence 33082, A
	39	18.2	72.8	429	4	US-09-925-065A-814695	Sequence 814695, A
	40	18.2	72.8	600	9	US-10-972-079-59759	Sequence 59759, A
	41	18.2	72.8	638	4	US-09-925-065A-450957	Sequence 450957, A
	42	18.2	72.8	660	4	US-09-925-065A-554692	Sequence 554692, A
	43	18.2	72.8	834	4	US-09-925-065A-681797	Sequence 681797, A
	44	18.2	72.8	919	5	US-10-027-632-121229	Sequence 121229, A
	45	18.2	72.8	919	5	US-10-027-632-121230	Sequence 121230, A
	46	18.2	72.8	919	6	US-10-027-632-121229	Sequence 121229, A
	47	18.2	72.8	919	6	US-10-027-632-121230	Sequence 121230, A
C	48	18.2	72.8	1136	5	US-10-027-632-117480	Sequence 117480, A
C	49	18.2	72.8	1136	6	US-10-027-632-117480	Sequence 117480, A
C	50	18.2	72.8	1175	7	US-10-424-599-42397	Sequence 42397, A
C	51	18.2	72.8	1985	5	US-10-027-632-97548	Sequence 97548, A
C	52	18.2	72.8	1985	5	US-10-027-632-99669	Sequence 99669, A
C	53	18.2	72.8	1985	6	US-10-027-632-97548	Sequence 97548, A
C	54	18.2	72.8	1985	6	US-10-027-632-99669	Sequence 99669, A
C	55	18.2	72.8	2906	5	US-10-027-632-249510	Sequence 249510, A
C	56	18.2	72.8	2906	6	US-10-027-632-249510	Sequence 249510, A
C	57	18.2	72.8	11241	7	US-10-322-281-656	Sequence 656, App
C	58	17.8	71.2	429	4	US-09-925-065A-277450	Sequence 277450, A
C	59	17.8	71.2	564	4	US-09-925-065A-781841	Sequence 781841, A
C	60	17.8	71.2	760	9	US-10-756-149-3366	Sequence 3366, Ap
C	61	17.8	71.2	2281	4	US-09-925-065A-722928	Sequence 722928, A
C	62	17.8	71.2	2281	4	US-09-925-065A-722929	Sequence 722929, A
C	63	17.8	71.2	1163020	7	US-10-398-221-10	Sequence 10, Appl
C	64	17.8	71.2	301208	7	US-10-398-221-2058	Sequence 2058, Ap
C	65	17.6	70.4	436	4	US-09-925-065A-335284	Sequence 335284, A
C	66	17.6	70.4	465	4	US-09-925-065A-478124	Sequence 478124, A
C	67	17.6	70.4	465	4	US-09-925-065A-478125	Sequence 478125, A
C	68	17.6	70.4	490	4	US-09-925-065A-601469	Sequence 601469, A
C	69	17.6	70.4	533	9	US-10-437-963-99308	Sequence 99308, A
C	70	17.6	70.4	600	9	US-10-972-079-34352	Sequence 34352, A
C	71	17.6	70.4	647	5	US-10-027-632-230481	Sequence 230481, A
C	72	17.6	70.4	647	6	US-10-027-632-230481	Sequence 230481, A
C	73	17.6	70.4	671	4	US-09-925-065A-770924	Sequence 770924, A
C	74	17.6	70.4	851	7	US-10-437-963-14931	Sequence 14931, A
C	75	17.6	70.4	925	6	US-10-369-493-32578	Sequence 32578, A
C	76	17.6	70.4	1303	7	US-10-425-114-4266	Sequence 4266, Ap
C	77	17.6	70.4	1368	10	US-11-097-143-14957	Sequence 14957, A
C	78	17.6	70.4	1563	3	US-09-160-116-1	Sequence 1, Appl
C	79	17.6	70.4	1590	4	US-09-925-065A-53317	Sequence 53317, A
C	80	17.6	70.4	1590	4	US-09-925-065A-53318	Sequence 53318, A
C	81	17.6	70.4	1660	7	US-10-437-963-14930	Sequence 14930, A
C	82	17.6	70.4	1867	3	US-09-879-957-13	Sequence 13, Appl
C	83	17.6	70.4	1867	7	US-10-807-856-13	Sequence 13, Appl
C	84	17.6	70.4	1972	7	US-10-437-963-14927	Sequence 14927, A
C	85	17.6	70.4	5010	10	US-11-097-143-14956	Sequence 14956, A
C	86	17.6	70.4	5904	10	US-11-097-143-25043	Sequence 25043, A
C	87	17.6	70.4	8685	10	US-11-097-143-25042	Sequence 25042, A
C	88	17.6	70.4	90468	6	US-10-085-117-160	Sequence 160, App
C	89	17.6	70.4	99232	8	US-10-331-053-16	Sequence 16, Appl
C	90	17.6	70.4	99588	5	US-10-087-192-1780	Sequence 1780, Ap
C	91	17.6	70.4	227968	8	US-10-723-860-1357	Sequence 1357, Ap
C	92	17.4	69.6	429	4	US-09-925-065A-277451	Sequence 277451, A
C	93	17.4	69.6	591	4	US-09-925-065A-659385	Sequence 659385, A
C	94	17.4	69.6	591	4	US-09-925-065A-659386	Sequence 659386, A
C	95	17.4	69.6	609	4	US-09-925-065A-355809	Sequence 355809, A
C	96	17.4	69.6	616	5	US-10-027-632-271180	Sequence 271180, A





```
; FILE REFERENCE: 03235/100J795-US4
; CURRENT APPLICATION NUMBER: US/11/099,266
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US 10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 74868
; TYPE: DNA
; ORGANISM: Mus musculus
US-11-099-266-67
```

```
Query Match 100.0%; Score 25; DB 10; Length 74868;
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAATCCTGTAGGCTTTTCAGTG 25
|||||
DB 33482 AAAAAAATCCTGTAGGCTTTTCAGTG 33458
```

## RESULT 4

```
US-10-719-900-19
; Sequence 19, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 19
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-19
```

```
Query Match 93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.24; 1; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAATCCTGTAGGCTTTTCAGTG 25
|||||
DB 1 AAAAAAATCCTGTAGGCTTTTCAGTG 25
```

## RESULT 5

```
US-10-425-115-33079/c
; Sequence 33079, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 33079
; LENGTH: 334
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_130178C.1
US-10-425-115-33079
```

```
Query Match 79.2%; Score 19.8; DB 8; Length 334;
Best Local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAATCCTGTAGGCTTTTCAG 23
|||||
DB 101 AAAAAAATCCTGTAGGCTGTGAG 79
```

## RESULT 6

```
US-10-788-792-104/c
; Sequence 104, Application US/10788792
; Publication No. US20040191819A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eyeleight, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 104
; LENGTH: 2243
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102)..(102)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (156)..(156)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (493)..(493)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (498)..(498)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (506)..(506)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (535)..(535)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (539)..(539)
```



```
; NAME/KEY: misc_feature
; LOCATION: (1890)..(1890)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature

Query Match          79.2%; Score 19.8; DB 8; Length 2243;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 2095 AAAAAATTCCTGTAGGTTTTCANT 2072

RESULT 7
US-10-242-535A-7089/c
; Sequence 7089, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7089
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-7089

Query Match          76.8%; Score 19.2; DB 7; Length 258;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 136 AAAAAATTCCTGTAGGTTTTCATT 113

RESULT 8
US-10-085-783A-7089/c
; Sequence 7089, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7089
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-7089

; ORGANISM: Human
US-10-085-783A-7089

Query Match          76.8%; Score 19.2; DB 7; Length 258;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 136 AAAAAATTCCTGTAGGTTTTCATT 113

RESULT 9
US-10-242-535A-38667/c
; Sequence 38667, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38667
; LENGTH: 289
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-38667

Query Match          76.8%; Score 19.2; DB 7; Length 289;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 166 AAAAAATTCCTGTAGGTTTTCATT 143

RESULT 10
US-10-085-783A-38667/c
; Sequence 38667, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38667
; LENGTH: 289
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-38667
```

```
Query Match          76.8%; Score 19.2; DB 7; Length 289;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 166 AAAAAATTCCTGTAGGTTTCATT 143

RESULT 11
US-10-125-968-712/c
; Sequence 712, Application US/10125968
; Publication No. US20030215805A1
; GENERAL INFORMATION:
; APPLICANT: Pallio, James
; APPLICANT: Palermo, Adam
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Elias, Josh
; APPLICANT: Mertens, Maureen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-032
; CURRENT APPLICATION NUMBER: US/10/125,968
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,163
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 1417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 712
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 290..291, 295, 297, 302, 308, 319, 340, 395, 408, 416, 423,
; LOCATION: 424
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-712

Query Match          76.8%; Score 19.2; DB 6; Length 439;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 166 AAAAAATTCCTGTAGGTTTCATT 143

RESULT 12
US-10-027-632-199786/c
; Sequence 199786, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199786
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199786

Query Match          76.8%; Score 19.2; DB 6; Length 546;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 339 AAAAAATTCCTGTAGGTTTCATT 316

RESULT 14
US-09-864-761-13669/c
; Sequence 13669, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Shaaron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199786
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199786

Query Match          76.8%; Score 19.2; DB 5; Length 546;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 339 AAAAAATTCCTGTAGGTTTCATT 316

RESULT 13
US-10-027-632-199786/c
; Sequence 199786, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199786
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199786

Query Match          76.8%; Score 19.2; DB 6; Length 546;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAGT 24
   ||||| ||||| ||||| ||||| |||||
Db 339 AAAAAATTCCTGTAGGTTTCATT 316

RESULT 14
US-09-864-761-13669/c
; Sequence 13669, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Shaaron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
```

; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13669  
; LENGTH: 562  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP000356.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
US-09-864-761-13669

Query Match 76.8%; Score 19.2; DB 3; Length 562;  
Best Local Similarity 87.5%; Pred. No. 45;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAATCCTCTAGGCTTTTCAGTG 25  
Db 29 AAAAATCCTCTAGGCTTTTCATTG 6

RESULT 15  
US-09-925-065A-81129/c  
; Sequence 81129, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81129  
; LENGTH: 693  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-81129

Query Match 76.8%; Score 19.2; DB 4; Length 693;  
Best Local Similarity 87.5%; Pred. No. 46;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAATCCTCTAGGCTTTTCAGTG 25  
Db 199 AAAAATCCTCTAGGCTTTTCAGTG 176

RESULT 16  
US-10-956-157-6011/c  
; Sequence 6011, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6011  
; LENGTH: 1400  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-6011

Query Match 76.8%; Score 19.2; DB 9; Length 1400;  
Best Local Similarity 87.5%; Pred. No. 51;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAATCCTCTAGGCTTTTCAGT 24  
Db 1276 AAAAATCCTCTAGGCTTTTCATT 1253

RESULT 17  
US-10-108-260A-731/c  
; Sequence 731, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 731  
; LENGTH: 2527  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-108-260A-731

Query Match 76.8%; Score 19.2; DB 6; Length 2527;

Best Local Similarity 87.5%; Pred. No. 55;  
Matches 21; Conservative 0; Mismatches

3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 2405 AAAAAAATCCTGTAGGTTTCATT 2382  
||||| ||||| ||||| ||||| |||||

RESULT 18  
US-10-146-473-6/c

; Sequence 6, Application US/10146473  
; Publication No. US2003010888A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew  
; APPLICANT: Gout, Ivan  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Gure, Ali  
; APPLICANT: Chen, Yao-Tsung  
; APPLICANT: Old, Lloyd  
; TITLE OF INVENTION: Breast Cancer Antigens  
; FILE REFERENCE: L00461/70130(JRV)  
; CURRENT APPLICATION NUMBER: US/10/146,473  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/291,150  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 2587  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-146-473-6

Query Match 76.8%; Score 19.2; DB 5; Length 2587;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 2463 AAAAAAATCCTGTAGGTTTCATT 2440  
||||| ||||| ||||| ||||| |||||

RESULT 19  
US-10-717-597-125/c

; Sequence 125, Application US/10717597  
; Publication No. US20040110221A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael E.  
; APPLICANT: Twine, Natalie C.  
; APPLICANT: Dörner, Andrew J.  
; APPLICANT: Trepicchio, William L.  
; APPLICANT: Slonim, Donna K.  
; APPLICANT: Stover, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS  
; FILE REFERENCE: AM101080L  
; CURRENT APPLICATION NUMBER: US/10/717,597  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: US 60/459,782  
; PRIOR FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: US 60/427,982  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 4904  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 125  
; LENGTH: 2587  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-717-597-125

Query Match 76.8%; Score 19.2; DB 7; Length 2587;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 2463 AAAAAAATCCTGTAGGTTTCATT 2440  
||||| ||||| ||||| ||||| |||||

RESULT 20

US-10-956-157-776/c  
; Sequence 776, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 776  
; LENGTH: 2587  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-776

Query Match 76.8%; Score 19.2; DB 9; Length 2587;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 2463 AAAAAAATCCTGTAGGTTTCATT 2440  
||||| ||||| ||||| ||||| |||||

RESULT 21

US-10-981-277-8/c  
; Sequence 8, Application US/10981277  
; Publication No. US20050181389A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Lisa  
; APPLICANT: Davis, Lisa  
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification  
; FILE REFERENCE: 03-968-US  
; CURRENT APPLICATION NUMBER: US/10/981,277  
; CURRENT FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/516,817  
; PRIOR FILING DATE: 2003-11-03  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 8  
; LENGTH: 2587  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-981-277-8

Query Match 76.8%; Score 19.2; DB 9; Length 2587;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGTTTCAGT 24  
||||| ||||| ||||| ||||| |||||  
DB 2463 AAAAAAATCCTGTAGGTTTCATT 2440  
||||| ||||| ||||| ||||| |||||

RESULT 22

US-10-252-157-57/c  
; Sequence 57, Application US/10252157  
; Publication No. US20030190840A1  
; GENERAL INFORMATION:  
; APPLICANT: Faris, Mary  
; APPLICANT: Pearson, Cecelia I.  
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER



```
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 57
; LENGTH: 3886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 314831.5
;
;
; NAME/KEY: unsure
; LOCATION: 226-288, 3667-3774
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-57

Query Match          76.8%; Score 19.2; DB 6; Length 3886;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
Db 2445 AAAAAATTCCTGTAGGTTTCATT 2422

RESULT 23
US-09-814-353-20129/c
; Sequence 20129, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20129
; LENGTH: 5569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 5569
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20129

Query Match          76.8%; Score 19.2; DB 3; Length 5569;
Best Local Similarity 87.5%; Pred. No. 62;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
Db 1 AAAAAATTCCTGTAGGTTTCATT 2422

FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 57
; LENGTH: 3886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 314831.5
;
;
; NAME/KEY: unsure
; LOCATION: 226-288, 3667-3774
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-57

Query Match          76.8%; Score 19.2; DB 6; Length 3886;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
Db 2445 AAAAAATTCCTGTAGGTTTCATT 2422

RESULT 24
US-10-357-930-24829/c
; Sequence 24829, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endegre, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24829
; LENGTH: 5570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 5570
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-24829

Query Match          76.8%; Score 19.2; DB 8; Length 5570;
Best Local Similarity 87.5%; Pred. No. 62;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGT 24
Db 317 AAAAAATTCCTGTAGGTTTCATT 294

RESULT 25
US-10-981-277-36
; Sequence 36, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 150525
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: n = A,T,C or G
US-10-981-277-36
```

```
Query Match          76.8%; Score 19.2; DB 9; Length 150525;
Best Local Similarity 87.5%; Pred. NO. 99;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

**Qy** 1 AAAAAA C T C C T G T A G G C T T T C A G T 24  
||| ||| ||| ||| ||| ||| ||| |||  
**Db** 92490 AAAAAA T T C C T G T A G G T T T T C A T T 92513

```

RESULT 26
US-10-984-277-35
; Sequence 35, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 35
; LENGTH: 166536
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-984-277-35

```

Query Match 76.8%; Score 19.2; DB 9; Length 166536;  
Best Local Similarity 87.5%; Pred. No. 1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 1 AAAAACTCCTGTAGGCTTTCAGT 24  
||||| ||||| ||||| |||||  
**D6** 26151 AAAAAATTCCTGTAGGTTTTTCATT 26174

```

RESULT 27
US-10-027-632-129877
; Sequence 129877, Application US/10027632
; Publication NO. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129877
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-129877

```

Query Match 74.4%; Score 18.6; DB 5; Length 594;

Best Local Similarity 84.0%; Pred. No. 89;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCAGTG 25  
||| ||| ||| ||| ||| ||| ||| |||  
pB 346 AAAAATCTCTCTGGGCTTCTGTG 370

```

RESULT 28
US-10-027-632-129878
; Sequence 129878, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129878
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-129878

```

Query Match	74.4%;	Score 18.6;	DB 5;	Length 594;
Best Local Similarity	84.0%;	Pred. No. 89;		
Matches	21.	Conservative	0:	Mismatches
			4:	Indels
			0:	Gaps

Qy 1 AAAAACTCCTGTAGGCTTTCAGTG 25  
||| ||| ||| ||| ||| ||| ||| |||  
Dh 245 AAAAAATCTCTCTCTGGCGTTTCTGTG 370

```

RESULT 29
US-10-027-632-129877
; Sequence 129877, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

```

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 129877  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-129877

Query Match 74.4%; Score 18.6; DB 6; Length 594;  
Best Local Similarity 84.0%; Pred. No. 89;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
DB 346 AAAAACTCTCTGGGCTTTTCGTG 370

## RESULT 30

US-10-027-632-129878  
; Sequence 129878, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 129878  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-129878

Query Match 74.4%; Score 18.6; DB 6; Length 594;  
Best Local Similarity 84.0%; Pred. No. 89;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
DB 346 AAAAACTCTCTGGGCTTTTCGTG 370

## RESULT 31

US-10-027-632-133274  
; Sequence 133274, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 133274  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-133274

Query Match 74.4%; Score 18.6; DB 5; Length 823;  
Best Local Similarity 84.0%; Pred. No. 94;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
DB 460 AGAAAACTCTGGAGGTTTCAGTG 484

## RESULT 32

US-10-027-632-133275  
; Sequence 133275, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 133275  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-133275

Query Match 74.4%; Score 18.6; DB 5; Length 823;  
Best Local Similarity 84.0%; Pred. No. 94;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
DB 460 AGAAAACTCTGGAGGTTTCAGTG 484

## RESULT 33

```
US-10-027-632-133274
; Sequence 133274, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133274
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-133274

Query Match      74.4%; Score 18.6; DB 6; Length 823;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

YQ 1 AAAAACTCTGTAGGCTTTTCAGTG 25
    ||||| ||||| ||||| ||||| |||||
Db 460 AGAAACTCTGTAGGCTTTTCAGTG 484

RESULT 34
US-10-027-632-133275
; Sequence 133275, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133275
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-133275

Query Match      74.4%; Score 18.6; DB 6; Length 823;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

YQ 1 AAAAACTCTGTAGGCTTTTCAGTG 25
    ||||| ||||| ||||| ||||| |||||
Db 460 AGAAACTCTGTAGGCTTTTCAGTG 484

RESULT 36
US-10-900-420-3
; Sequence 3, Application US/10900420
; Publication No. US20050003434A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; THEREOF
; FILE REFERENCE: CL001213DIV-II
; CURRENT APPLICATION NUMBER: US/10/900,420
; CURRENT FILING DATE: 2004-07-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 39982
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-900-420-3

Query Match      74.4%; Score 18.6; DB 8; Length 39982;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

YQ 1 AAAAACTCTGTAGGCTTTTCAGTG 25
    ||||| ||||| ||||| ||||| |||||
Db 14081 AAAGAAATCTGTGGCTTTCTGTG 14105

RESULT 37
US-10-425-115-172275/c
; Sequence 172275, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172275
; LENGTH: 317
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_88696C.1
US-10-425-115-172275

Query Match          72.8%; Score 18.2; DB 8; Length 317;
Best Local Similarity 87.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAG 23
   ||||| ||||| ||||| ||||| |||||
Db 38 AAAAACTGCTGTAGGCTGTGAG 16

RESULT 38
US-10-425-115-33082/c
; Sequence 33082, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 33082
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_130180C.1
US-10-425-115-33082

Query Match          72.8%; Score 18.2; DB 8; Length 351;
Best Local Similarity 87.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAG 23
   ||||| ||||| ||||| ||||| |||||
Db 102 AAAAACTCTGTAGGCTGTGAG 80

RESULT 39
US-09-925-065A-814695
; Sequence 814695, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
```

```
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 814695
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-814695

Query Match          72.8%; Score 18.2; DB 4; Length 429;
Best Local Similarity 87.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAATCTCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| ||||| |||||
Db 22 AAAAATCTCTAAAGGCTATCATGTG 44

RESULT 40
US-10-972-079-59759
; Sequence 59759, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59759
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894318382_1
US-10-972-079-59759

Query Match          72.8%; Score 18.2; DB 9; Length 600;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAATCTCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| ||||| |||||
Db 362 AAAAATCTCTGTAGGCTTTTCATG 384

RESULT 41
US-09-925-065A-450957
; Sequence 450957, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
```

```
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 450957
; LENGTH: 638
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-450957
```

```
Query Match 72.8%; Score 18.2; DB 4; Length 638;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAATCCTCTAGGCTTTTCAG 23
    ||||| ||||| ||||| |||||
Db 348 AAACAAATCCTATAGGCTTTTCAG 370
```

```
RESULT 42
US-09-925-065A-554692
; Sequence 554692, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 554692
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-554692
```

```
Query Match 72.8%; Score 18.2; DB 4; Length 660;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAATCCTCTAGGCTTTTCAG 23
    ||||| ||||| ||||| |||||
Db 264 AAACAAATCCTATAGGCTTTTCAG 286
```

```
RESULT 43
US-09-925-065A-681797
; Sequence 681797, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
```

```
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 681797
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-681797
```

```
Query Match 72.8%; Score 18.2; DB 4; Length 834;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AAAATCCTCTAGGCTTTTCAGTG 25
    ||||| ||||| ||||| |||||
Db 149 AAAATCCTCTAGGCTTTTCAGTG 171
```

```
RESULT 44
US-10-027-632-121229
; Sequence 121229, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121229
; LENGTH: 919
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-121229
```

```
Query Match 72.8%; Score 18.2; DB 5; Length 919;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AAAATCCTCTAGGCTTTTCAGTG 25
    ||||| ||||| ||||| |||||
Db 645 AAAGAGCTGTAGGCTTTTCAGTG 667
```

```
RESULT 45
US-10-027-632-121230
; Sequence 121230, Application US/10027632
; Publication No. US20020198371A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121230
; LENGTH: 919
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-121230
```

```

Query Match 72.8%; Score 18.2; DB 5; Length 919;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 3 AAAAATCCTGTAGGCTTTCAGTG 25
    ||||| ||||| ||||| |||||
Db 645 AAAAGAGCTGTAGGCTTTCAGTG 667
```

```

RESULT 46
US-10-027-632-121229
; Sequence 121229, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121229
; LENGTH: 919
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-121229
```

```

Query Match 72.8%; Score 18.2; DB 6; Length 919;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
```

```

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTCAGTG 25
    ||||| ||||| ||||| |||||
Db 645 AAAAGAGCTGTAGGCTTTCAGTG 667
```

```

RESULT 47
US-10-027-632-121230
; Sequence 121230, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121230
; LENGTH: 919
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-121230
```

```

Query Match 72.8%; Score 18.2; DB 6; Length 919;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 3 AAAAATCCTGTAGGCTTTCAGTG 25
    ||||| ||||| ||||| |||||
Db 645 AAAAGAGCTGTAGGCTTTCAGTG 667
```

```

RESULT 48
US-10-027-632-117480/C
; Sequence 117480, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
```



```
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117480
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1136)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-117480
```

```
Query Match 72.8%; Score 18.2; DB 5; Length 1136;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AAAACTCCTGTAGGCTTTTCAGTG 25
Db 573 AAAACTCCTTTATGTTTTCAGTG 551
```

```
RESULT 49
US-10-027-632-117480/c
; Sequence 117480, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117480
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1136)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-117480
```

```
Query Match 72.8%; Score 18.2; DB 6; Length 1136;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 AAAACTCCTGTAGGCTTTTCAGTG 25
Db 573 AAAACTCCTTTATGTTTTCAGTG 551
```

```
RESULT 50
US-10-424-599-42397/c
; Sequence 42397, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
```

```
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; PLANT GROWTH AND YIELD IMPROVEMENT
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 42397
; LENGTH: 1175
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138285C.1
US-10-424-599-42397
```

```
Query Match 72.8%; Score 18.2; DB 7; Length 1175;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 AAAACTCCTGTAGGCTTTTCAGT 24
Db 587 AAAACACCTATAGGTTTCAGT 565
```

```
Search completed: February 3, 2006, 15:43:52
Job time : 371.556 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-20

Perfect score: 25

Sequence: 1 aaaaactcctgtaggcttcagtg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New.\*

1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	18.6	74.4	4499	7	US-10-750-185-43501
C 2	18.6	74.4	4499	7	US-10-750-623-43501
C 3	17.6	70.4	568	8	US-11-128-061-1900
C 4	17.6	70.4	568	8	US-11-128-061-5542
C 5	17.6	70.4	568	8	US-11-128-049-1900
C 6	17.6	70.4	568	8	US-11-128-049-5542
C 7	17.2	68.8	6138	8	US-11-180-000-4
C 8	17.2	68.8	6141	8	US-11-180-000-3
C 9	17.2	68.8	7646	8	US-11-180-000-1
C 10	17.2	68.8	16570	8	US-11-180-000-2
C 11	17.2	68.8	68123	7	US-10-995-561-13348
C 12	17	68.0	201	8	US-11-124-368A-19365
C 13	17	68.0	201	8	US-11-124-368A-19366
C 14	17	68.0	1350	7	US-10-750-185-52644
C 15	17	68.0	1350	7	US-10-750-623-52644
C 16	17	68.0	1516	7	US-10-750-185-39342
C 17	17	68.0	1516	7	US-10-750-623-39342
C 18	17	68.0	1570	7	US-10-750-185-36386
C 19	17	68.0	1570	7	US-10-750-623-36386
C 20	17	68.0	1728	7	US-10-750-185-43815
C 21	17	68.0	1728	7	US-10-750-623-43815
C 22	17	68.0	1954	7	US-10-750-185-43339

1954	68.0	17	7	US-10-750-623-43339	Sequence 43339, A
2064	68.0	17	8	US-11-117-169-5	Sequence 5, Appli
2186	68.0	17	8	US-11-117-169-7	Sequence 7, Appli
2417	68.0	17	7	US-10-750-185-51243	Sequence 51243, A
2417	68.0	17	7	US-10-750-623-51243	Sequence 51243, A
2526	68.0	17	8	US-11-052-554A-486	Sequence 486, App
2843	68.0	17	7	US-10-750-185-64398	Sequence 64398, A
2843	68.0	17	7	US-10-750-623-64398	Sequence 64398, A
3068	68.0	17	8	US-11-117-169-14	Sequence 14, Appli
325	67.2	16.8	8	US-11-121-849-393885	Sequence 393885, A
747	67.2	16.8	7	US-10-750-185-52818	Sequence 52818, A
747	67.2	16.8	7	US-10-750-623-52818	Sequence 52818, A
600	66.4	16.6	7	US-10-750-185-4713	Sequence 4713, Ap
600	66.4	16.6	7	US-10-750-185-21331	Sequence 21331, A
600	66.4	16.6	7	US-10-750-623-4713	Sequence 4713, Ap
600	66.4	16.6	7	US-10-750-623-21331	Sequence 21331, A
781	66.4	16.6	7	US-10-750-185-26142	Sequence 26142, A
781	66.4	16.6	7	US-10-750-623-26142	Sequence 26142, A
914	66.4	16.6	7	US-10-750-185-52730	Sequence 52730, A
914	66.4	16.6	7	US-10-750-623-52730	Sequence 52730, A
1153	66.4	16.6	7	US-10-750-185-57290	Sequence 57290, A
1153	66.4	16.6	7	US-10-750-623-57290	Sequence 57290, A
1496	66.4	16.6	7	US-10-750-185-44378	Sequence 44378, A
1496	66.4	16.6	7	US-10-750-623-44378	Sequence 44378, A
1507	66.4	16.6	7	US-10-750-185-52605	Sequence 52605, A
1507	66.4	16.6	7	US-10-750-623-52605	Sequence 52605, A
1695	66.4	16.6	7	US-10-750-185-50030	Sequence 50030, A
1695	66.4	16.6	7	US-10-750-623-50030	Sequence 50030, A
2251	66.4	16.6	7	US-10-750-185-42351	Sequence 42351, A
2251	66.4	16.6	7	US-10-750-623-42351	Sequence 42351, A
4168	66.4	16.6	7	US-10-750-185-50907	Sequence 50907, A
4168	66.4	16.6	7	US-10-750-623-50907	Sequence 50907, A
83712	66.4	16.6	7	US-10-995-561-13366	Sequence 13366, A
178877	66.4	16.6	8	US-11-121-086-17	Sequence 17, Appli
188056	66.4	16.6	8	US-11-120-925-1	Sequence 1, Appli
969	64.8	16.2	7	US-10-750-185-34140	Sequence 34140, A
969	64.8	16.2	7	US-10-750-623-34140	Sequence 34140, A
1028	64.8	16.2	7	US-10-750-185-53981	Sequence 53981, A
1028	64.8	16.2	7	US-10-750-623-53981	Sequence 53981, A
1032	64.8	16.2	7	US-10-750-185-25195	Sequence 25195, A
1032	64.8	16.2	7	US-10-750-623-25195	Sequence 25195, A
1324	64.8	16.2	7	US-10-750-185-52953	Sequence 52953, A
1324	64.8	16.2	7	US-10-750-623-52953	Sequence 52953, A
1881	64.8	16.2	7	US-10-750-185-52504	Sequence 52504, A
1881	64.8	16.2	7	US-10-750-623-52504	Sequence 52504, A
2303	64.8	16.2	7	US-10-750-185-35931	Sequence 35931, A
2303	64.8	16.2	7	US-10-750-623-35931	Sequence 35931, A
2656	64.8	16.2	7	US-10-750-185-31466	Sequence 31466, A
2656	64.8	16.2	7	US-10-750-623-31466	Sequence 31466, A
2881	64.8	16.2	8	US-11-136-527-576	Sequence 576, App
2881	64.8	16.2	8	US-11-136-527-576	Sequence 576, App
17004	64.8	16.2	8	US-11-176-253-1	Sequence 1, Appli
17004	64.8	16.2	8	US-11-176-253-1	Sequence 1, Appli
611587	64.8	16.2	8	US-11-117-187-209	Sequence 209, Appli
1691140	64.8	16.2	8	US-11-091-018-1	Sequence 1, Appli
567	64.0	16.0	7	US-10-793-626-3271	Sequence 3271, Ap
884	64.0	16.0	7	US-10-750-185-58951	Sequence 58951, A
884	64.0	16.0	7	US-10-750-623-58951	Sequence 58951, A
1028	64.0	16.0	7	US-10-750-185-48397	Sequence 48397, A
1028	64.0	16.0	7	US-10-750-623-48397	Sequence 48397, A
1141	64.0	16.0	7	US-10-750-185-40232	Sequence 40232, A
1141	64.0	16.0	7	US-10-750-623-40232	Sequence 40232, A
1227	64.0	16.0	7	US-10-750-185-63598	Sequence 63598, A
1227	64.0	16.0	7	US-10-750-623-63598	Sequence 63598, A
1245	64.0	16.0	7	US-10-467-657-9	Sequence 9, Appli
1245	64.0	16.0	7	US-10-467-657-9	Sequence 9, Appli
1245	64.0	16.0	7	US-10-750-185-51455	Sequence 51455, A
1338	64.0	16.0	7	US-10-750-623-51455	Sequence 51455, A
1340	64.0	16.0	7	US-10-750-185-28818	Sequence 28818, A
1340	64.0	16.0	7	US-10-750-623-28818	Sequence 28818, A
1448	64.0	16.0	5	US-09-978-360A-26	Sequence 26, Appli
1448	64.0	16.0	5	US-09-978-360A-26	Sequence 26, Appli
1572	64.0	16.0	7	US-10-131-826A-373	Sequence 373, App
1572	64.0	16.0	7	US-10-131-826A-373	Sequence 373, App
1634	64.0	16.0	7	US-10-131-826A-181	Sequence 181, App
1634	64.0	16.0	7	US-10-131-826A-181	Sequence 181, App
1630	64.0	16.0	7	US-10-750-185-56730	Sequence 56730, A

96	16	64.0	1630	7	US-10-750-623-56730	Sequence 56730, A
97	16	64.0	1945	8	US-11-128-061-112	Sequence 112, App
98	16	64.0	1945	8	US-11-128-049-112	Sequence 112, App
99	16	64.0	2113	7	US-10-750-185-64322	Sequence 64322, A
100	16	64.0	2113	7	US-10-750-623-64322	Sequence 64322, A
101	16	64.0	2694	8	US-11-137-465-2	Sequence 2, Appli
102	16	64.0	2777	7	US-10-750-185-52831	Sequence 52831, A
103	16	64.0	2777	7	US-10-750-623-52831	Sequence 52831, A
104	16	64.0	2982	8	US-11-137-465-3	Sequence 3, Appli
105	16	64.0	3126	7	US-10-750-185-58454	Sequence 58454, A
106	16	64.0	3126	7	US-10-750-623-58454	Sequence 58454, A
107	16	64.0	3487	7	US-10-750-185-48264	Sequence 48264, A
108	16	64.0	3487	7	US-10-750-623-48264	Sequence 48264, A
109	16	64.0	3660	8	US-11-204-187-21	Sequence 21, Appl
110	16	64.0	3864	7	US-11-037-243-13	Sequence 13, Appl
111	16	64.0	3986	7	US-10-793-626-4014	Sequence 4014, Ap
112	16	64.0	10619	7	US-10-240-708-4	Sequence 4, Appli
113	16	64.0	11461	8	US-11-230-251-17	Sequence 17, Appl
114	16	64.0	17770	8	US-11-124-368A-2874	Sequence 2874, Ap
115	16	64.0	55763	7	US-10-972-766-1	Sequence 1, Appli
116	16	64.0	151152	7	US-10-775-168-243	Sequence 243, App
117	16	64.0	170995	8	US-11-121-086-35	Sequence 35, Appl
118	16	64.0	182190	8	US-11-121-086-102	Sequence 102, App
119	16	64.0	184000	8	US-11-121-086-37	Sequence 37, Appl
120	16	64.0	645179	7	US-10-995-561-13293	Sequence 13293, A
121	15.8	63.2	25	8	US-11-136-527-250500	Sequence 250500,
122	15.8	63.2	201	7	US-10-995-561-31525	Sequence 31525, A
123	15.8	63.2	201	8	US-11-124-368A-3346	Sequence 3346, Ap
124	15.8	63.2	201	8	US-11-124-368A-4124	Sequence 4124, Ap
125	15.8	63.2	570	8	US-11-136-527-1208	Sequence 1208, Ap
126	15.8	63.2	570	8	US-11-136-527-5304	Sequence 5304, Ap
127	15.8	63.2	600	7	US-10-750-185-944	Sequence 944, App
128	15.8	63.2	600	7	US-10-750-623-944	Sequence 944, App
129	15.8	63.2	600	8	US-11-128-061-4421	Sequence 4421, Ap
130	15.8	63.2	600	8	US-11-128-049-4421	Sequence 4421, Ap
131	15.8	63.2	783	7	US-10-750-185-33218	Sequence 33218, A
132	15.8	63.2	783	7	US-10-750-623-33218	Sequence 33218, A
133	15.8	63.2	1262	8	US-11-128-061-779	Sequence 779, App
134	15.8	63.2	1262	8	US-11-128-049-779	Sequence 779, App
135	15.8	63.2	1465	7	US-10-750-185-34071	Sequence 34071, A
136	15.8	63.2	1465	7	US-10-750-623-34071	Sequence 34071, A
137	15.8	63.2	1955	7	US-10-750-185-32634	Sequence 32634, A
138	15.8	63.2	1955	7	US-10-750-623-32634	Sequence 32634, A
139	15.8	63.2	4290	8	US-11-136-527-3034	Sequence 3034, Ap
140	15.8	63.2	101001	7	US-10-995-561-13255	Sequence 13255, A
141	15.8	63.2	156544	8	US-11-121-086-81	Sequence 81, Appl
142	15.6	62.4	22	7	US-10-310-914A-298511	Sequence 298511,
143	15.6	62.4	459	7	US-10-467-657-7673	Sequence 7673, Ap
144	15.6	62.4	459	7	US-10-467-657-8431	Sequence 8431, Ap
145	15.6	62.4	467	8	US-11-000-688-769	Sequence 769, App
146	15.6	62.4	505	8	US-11-128-061-1581	Sequence 1581, Ap
147	15.6	62.4	505	8	US-11-128-061-5223	Sequence 5223, Ap
148	15.6	62.4	505	8	US-11-128-049-1581	Sequence 1581, Ap
149	15.6	62.4	505	8	US-11-128-049-5223	Sequence 5223, Ap
150	15.6	62.4	838	7	US-10-750-185-51330	Sequence 51330, A

ALIGNMENTS

RESULT 1

US-10-750-185-43501/c

Sequence 43501, Application US/10750185

Publication No. US20050287531A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

RESULT 2

US-10-750-623-43501/c

Sequence 43501, Application US/10750623

Publication No. US20050287531A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

RESULT 3

US-11-128-061-1900

Sequence 1900, Application US/11128061

Publication No. US20060003958A1

GENERAL INFORMATION:

APPLICANT: Melville, Mark W.

APPLICANT: Charlebois, Timothy S.

APPLICANT: Mounts, William M.

APPLICANT: Hann, Louane E.

APPLICANT: Sinacore, Martin S.

APPLICANT: Leonard, Mark W.

APPLICANT: Brown, Eugene L.

APPLICANT: Miller, Christopher P.

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS

TITLE OF INVENTION: TO MONITOR GENE EXPRESSION

FILE REFERENCE: 01997.027701

FILE REFERENCE: MM1100-2	
CURRENT APPLICATION NUMBER: US/10/750,185	
CURRENT FILING DATE: 2003-12-31	
PRIOR APPLICATION NUMBER: US 60/437,482	
PRIOR FILING DATE: 2002-12-31	
NUMBER OF SEQ ID NOS: 64922	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 43501	
LENGTH: 4499	
TYPE: DNA	
ORGANISM: Bovine	19866880486171
US-10-750-185-43501	
Query Match	74.4%; Score 18.6; DB 7; Length 4499;
Best Local Similarity	84.0%; Pred. No. 14;
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	1 AAAAAAACCCTCTAGGCTTTTCAGTG 25
Db	3022 AAAAAAACCCTCTAGGCTTTTCAGTG 2998
RESULT 2	
US-10-750-623-43501/c	
Sequence 43501, Application US/10750623	
Publication No. US20050287531A1	
GENERAL INFORMATION:	
APPLICANT: MMI GENOMICS, INC.	
APPLICANT: DENISE, Sue K.	
APPLICANT: KERR, Richard	
APPLICANT: ROSENFELD, David	
APPLICANT: HOLM, Tom	
APPLICANT: BATES, Stephen	
APPLICANT: FANTIN, Dennis	
TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS	
FILE REFERENCE: MM1100-1	
CURRENT APPLICATION NUMBER: US/10/750,623	
CURRENT FILING DATE: 2003-12-31	
PRIOR APPLICATION NUMBER: US 60/437,482	
PRIOR FILING DATE: 2002-12-31	
NUMBER OF SEQ ID NOS: 64922	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 43501	
LENGTH: 4499	
TYPE: DNA	
ORGANISM: Bovine	19866880486171
US-10-750-623-43501	
Query Match	74.4%; Score 18.6; DB 7; Length 4499;
Best Local Similarity	84.0%; Pred. No. 14;
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	1 AAAAAAACCCTCTAGGCTTTTCAGTG 25
Db	3022 AAAAAAACCCTCTAGGCTTTTCAGTG 2998
RESULT 3	
US-11-128-061-1900	
Sequence 1900, Application US/11128061	
Publication No. US20060003958A1	
GENERAL INFORMATION:	
APPLICANT: Melville, Mark W.	
APPLICANT: Charlebois, Timothy S.	
APPLICANT: Mounts, William M.	
APPLICANT: Hann, Louane E.	
APPLICANT: Sinacore, Martin S.	
APPLICANT: Leonard, Mark W.	
APPLICANT: Brown, Eugene L.	
APPLICANT: Miller, Christopher P.	
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS	
TITLE OF INVENTION: TO MONITOR GENE EXPRESSION	
FILE REFERENCE: 01997.027701	



; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: EP 10116044.7  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: EP 01104171.2  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 6138  
; TYPE: DNA  
; ORGANISM: Human cytomegalovirus  
US-11-180-000-4

Query Match 68.8%; Score 17.2; DB 8; Length 6138;  
Best Local Similarity 86.4%; Pred. No. 71;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
||||| ||||| ||||| ||||| |||||  
DB 1781 AAAAACTCTGTGGCGTTCA 1760

## RESULT 8

US-11-180-000-3/c  
; Sequence 3, Application US/11180000  
; Publication No. US20060019369A1  
; GENERAL INFORMATION:

; APPLICANT: Hahn, Gabriele  
; TITLE OF INVENTION: Recombinant vector containing infectious human cytomegalovirus ge  
; TITLE OF INVENTION: with preserved wild-type characteristics of clinical isolates  
; FILE REFERENCE: 2923-212  
; CURRENT APPLICATION NUMBER: US/11/180,000  
; CURRENT FILING DATE: 2005-07-13  
; PRIOR APPLICATION NUMBER: US/10/275,287  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/01867  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: EP 10116044.7  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: EP 01104171.2  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 6141  
; TYPE: DNA  
; ORGANISM: Human cytomegalovirus  
US-11-180-000-3

Query Match 68.8%; Score 17.2; DB 8; Length 6141;  
Best Local Similarity 86.4%; Pred. No. 71;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
||||| ||||| ||||| ||||| |||||  
DB 1781 AAAAACTCTGTGGCGTTCA 1760

## RESULT 9

US-11-180-000-1/c  
; Sequence 1, Application US/11180000  
; Publication No. US20060019369A1  
; GENERAL INFORMATION:

; APPLICANT: Hahn, Gabriele  
; TITLE OF INVENTION: Recombinant vector containing infectious human cytomegalovirus ge  
; TITLE OF INVENTION: with preserved wild-type characteristics of clinical isolates  
; FILE REFERENCE: 2923-212  
; CURRENT APPLICATION NUMBER: US/11/180,000  
; CURRENT FILING DATE: 2005-07-13  
; PRIOR APPLICATION NUMBER: US/10/275,287  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/01867  
; PRIOR FILING DATE: 2002-02-21

; PRIOR APPLICATION NUMBER: EP 10116044.7  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: EP 01104171.2  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 7646  
; TYPE: DNA  
; ORGANISM: Human cytomegalovirus  
US-11-180-000-1

Query Match 68.8%; Score 17.2; DB 8; Length 7646;  
Best Local Similarity 86.4%; Pred. No. 75;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
||||| ||||| ||||| ||||| |||||  
DB 3273 AAAAACTCTGTGGCGTTCA 3252

## RESULT 10

US-11-180-000-2/c  
; Sequence 2, Application US/11180000  
; Publication No. US20060019369A1  
; GENERAL INFORMATION:

; APPLICANT: Hahn, Gabriele  
; TITLE OF INVENTION: Recombinant vector containing infectious human cytomegalovirus g  
; TITLE OF INVENTION: with preserved wild-type characteristics of clinical isolates  
; FILE REFERENCE: 2923-212  
; CURRENT APPLICATION NUMBER: US/11/180,000  
; CURRENT FILING DATE: 2005-07-13  
; PRIOR APPLICATION NUMBER: US/10/275,287  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: PCT/EP02/01867  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: EP 10116044.7  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: EP 01104171.2  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 16570  
; TYPE: DNA  
; ORGANISM: Human cytomegalovirus  
US-11-180-000-2

Query Match 68.8%; Score 17.2; DB 8; Length 16570;  
Best Local Similarity 86.4%; Pred. No. 88;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTCA 22  
||||| ||||| ||||| ||||| |||||  
DB 12206 AAAAACTCTGTGGCGTTCA 12185

## RESULT 11

US-10-995-561-13348  
; Sequence 13348, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13348  
; LENGTH: 68123







US-10-750-185-43815  
; Sequence 43815, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-2

; CURRENT APPLICATION NUMBER: US/10/750,185

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 43815

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Bovine 19866880856399

US-10-750-185-43815

Query Match 68.0%; Score 17; DB 7; Length 1728;  
Best Local Similarity 80.0%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25

Db 44 AAAAAATTCAGTATCTTTTCAGTG 68

RESULT 21

US-10-750-623-43815

; Sequence 43815, Application US/10750623

; Publication No. US20050287531A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-1

; CURRENT APPLICATION NUMBER: US/10/750,623

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 43815

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Bovine 19866880856399

US-10-750-623-43815

Query Match 68.0%; Score 17; DB 7; Length 1728;  
Best Local Similarity 80.0%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25

Db 44 AAAAAATTCAGTATCTTTTCAGTG 68

RESULT 22

US-10-750-185-43339/c

; Sequence 43339, Application US/10750185

; Publication No. US20050260603A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-2

; CURRENT APPLICATION NUMBER: US/10/750,185

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 43339

; LENGTH: 1954

; TYPE: DNA

; ORGANISM: Bovine 19866880591323

US-10-750-185-43339

Query Match 68.0%; Score 17; DB 7; Length 1954;  
Best Local Similarity 80.0%; Pred. No. 71;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25

Db 1748 AAAAAATGCTAAGGATTTTCAGTG 1724

RESULT 23

US-10-750-623-43339/c

; Sequence 43339, Application US/10750623

; Publication No. US20050287531A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-1

; CURRENT APPLICATION NUMBER: US/10/750,623

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 43339

; LENGTH: 1954

; TYPE: DNA

; ORGANISM: Bovine 19866880591323

US-10-750-623-43339

Query Match 68.0%; Score 17; DB 7; Length 1954;  
Best Local Similarity 80.0%; Pred. No. 71;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25

Db 1748 AAAAAATGCTAAGGATTTTCAGTG 1724

RESULT 24

US-11-117-169-5/c

; Sequence 5, Application US/11117169

; Publication No. US2005026609A1

; GENERAL INFORMATION:

; APPLICANT: University of South Florida

; APPLICANT: Mohapatra, Shyam

```
/ TITLE OF INVENTION: Nanogene Therapy For Cell Proliferation Disorders
/ FILE REFERENCE: USF-216PTCZ
/ CURRENT APPLICATION NUMBER: US/11/117,169
/ CURRENT FILING DATE: 2005-04-27
/ PRIOR APPLICATION NUMBER: 60/565,756
/ PRIOR FILING DATE: 2004-04-27
/ NUMBER OF SEQ ID NOS: 31
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 5
/ LENGTH: 2064
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(2064)
/ PUBLICATION INFORMATION:
/ AUTHORS: Hovnanian, A., et al.
/ TITLE: The human 2', 5'-oligoadenylate synthetase locus is composed of
/ TITLE: three distinct genes
/ JOURNAL: Genomics
/ VOLUME: 52
/ ISSUE: 3
/ PAGES: 267-277
/ DATE: 1998
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ PUBLICATION INFORMATION:
/ AUTHORS: Marie, I. and Hovanessian, A.G.
/ TITLE: The 69-kDa 2-5A synthetase is composed of two homologous and
/ TITLE: adjacent functional domains
/ JOURNAL: J. Biol. Chem.
/ VOLUME: 267
/ ISSUE: 14
/ PAGES: 9933-9939
/ DATE: 1992
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ PUBLICATION INFORMATION:
/ AUTHORS: Marie, I., et al.
/ TITLE: Differential expression and distinct structure of 69- and 100-kDa
/ TITLE: forms of 2-5A synthetase
/ JOURNAL: J. Biol. Chem.
/ VOLUME: 265
/ ISSUE: 30
/ PAGES: 18601-18607
/ DATE: 1990
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ PUBLICATION INFORMATION:
/ AUTHORS: Marie, I., et al.
/ TITLE: Preparation and characterization of polyclonal antibodies
/ TITLE: specific for the 69 and 100 k-dalton forms of human 2-5A
/ JOURNAL: Biochem. Biophys. Res. Commun.
/ VOLUME: 160
/ ISSUE: 2
/ PAGES: 580-587
/ DATE: 1989
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ PUBLICATION INFORMATION:
/ AUTHORS: Hovanessian, A.G., et al.
/ TITLE: Characterization of 69- and 100-kDa forms of 2-5A-synthetase from
/ TITLE: interferon-treated human cells
/ JOURNAL: J. Biol. Chem.
/ VOLUME: 263
/ ISSUE: 10
/ PAGES: 4959
/ DATE: 1988
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ PUBLICATION INFORMATION:
/ AUTHORS: Hovanessian, A.G., et al.
```

```
/ TITLE: Identification of 69-kd and 100-kd forms of 2-5A synthetase
/ JOURNAL: EMBO J.
/ VOLUME: 6
/ ISSUE: 5
/ PAGES: 1273-1280
/ DATE: 1987
/ DATABASE ACCESSION NUMBER: NCBI/NM_002535
/ DATABASE ENTRY DATE: 2003-04-03
/ US-11-117-169-5

Query Match      68.0%; Score 17; DB 8; Length 2064;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||| ||| ||| ||| |||
Db 567 AAAAAAATCCTGTGAGTTTCAGTG 543

RESULT 25
US-11-117-169-7/c
/ Sequence 7, Application US/11117169
/ Publication No. US20050266093A1
/ GENERAL INFORMATION:
/ APPLICANT: University of South Florida
/ APPLICANT: Mohapatra, Shyam
/ TITLE OF INVENTION: Nanogene Therapy For Cell Proliferation Disorders
/ FILE REFERENCE: USF-216PTCZ
/ CURRENT APPLICATION NUMBER: US/11/117,169
/ CURRENT FILING DATE: 2005-04-27
/ PRIOR APPLICATION NUMBER: 60/565,756
/ PRIOR FILING DATE: 2004-04-27
/ NUMBER OF SEQ ID NOS: 31
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 7
/ LENGTH: 2186
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ PUBLICATION INFORMATION:
/ AUTHORS: Marie, I. and Hovanessian, A.G.
/ TITLE: The 69-kDa 2-5A synthetase is composed of two homologous and
/ TITLE: adjacent functional domains
/ JOURNAL: J. Biol. Chem.
/ VOLUME: 267
/ ISSUE: 14
/ PAGES: 9933-9939
/ DATE: 1992
/ DATABASE ACCESSION NUMBER: (unknown)
/ DATABASE ENTRY DATE: 2003-04-03
/ US-11-117-169-7

Query Match      68.0%; Score 17; DB 8; Length 2186;
Best Local Similarity 80.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||| ||| ||| ||| |||
Db 567 AAAAAAATCCTGTGAGTTTCAGTG 543

RESULT 26
US-10-750-185-51243/c
/ Sequence 51243, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51243  
; LENGTH: 2417  
; TYPE: DNA  
; ORGANISM: Bovine 19866880878023  
US-10-750-185-51243

Query Match 68.0%; Score 17; DB 7; Length 2417;  
Best Local Similarity 80.0%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
Db 519 AAAAAACAGGAGTAGGTTTTCAGTG 495

RESULT 27  
US-10-750-623-51243/c  
; Sequence 51243, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51243  
; LENGTH: 2417  
; TYPE: DNA  
; ORGANISM: Bovine 19866880878023  
US-10-750-623-51243

Query Match 68.0%; Score 17; DB 7; Length 2417;  
Best Local Similarity 80.0%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
Db 519 AAAAAACAGGAGTAGGTTTTCAGTG 495

RESULT 28  
US-11-052-554A-486  
; Sequence 486, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 486  
; LENGTH: 2526  
; TYPE: DNA  
; ORGANISM: Helicobacter pylori J99  
US-11-052-554A-486

Query Match 68.0%; Score 17; DB 8; Length 2526;  
Best Local Similarity 80.0%; Pred. No. 75;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
Db 7 AAAAAAGTTCTGTGGCTTTTAACTG 31

RESULT 29  
US-10-750-185-64398  
; Sequence 64398, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64398  
; LENGTH: 2843  
; TYPE: DNA  
; ORGANISM: Bovine 19866880746758  
US-10-750-185-64398

Query Match 68.0%; Score 17; DB 7; Length 2843;  
Best Local Similarity 80.0%; Pred. No. 77;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAGTG 25  
||||| ||||| ||||| ||||| |||||  
Db 717 AAAAACTCTTTAAGCATTTCAGAG 741

RESULT 30  
US-10-750-623-64398  
; Sequence 64398, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64398  
; LENGTH: 2843



```
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52818
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Bovine 19866881647514
US-10-750-185-52818

Query Match      67.2%; Score 16.8; DB 7; Length 747;
Best Local Similarity 90.0%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 AAACTCTGTAGGCTTTTCAG 23
Db      564 AAATCTGTAGTCTTTTCAG 545

RESULT 34
US-10-750-623-52818/c
; Sequence 52818, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52818
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Bovine 19866881647514
US-10-750-623-52818

Query Match      67.2%; Score 16.8; DB 7; Length 747;
Best Local Similarity 90.0%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 AAACTCTGTAGGCTTTTCAG 23
Db      564 AAATCTGTAGTCTTTTCAG 545

RESULT 35
US-10-750-185-4713/c
; Sequence 4713, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

```
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 4713
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT04380
US-10-750-185-4713

Query Match      66.4%; Score 16.6; DB 7; Length 600;
Best Local Similarity 82.6%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      3 AAAACTCCTGTAGGCTTTTCAGTG 25
Db      129 AATACTCCTGGGGCTTCCAGTG 107

RESULT 36
US-10-750-185-21331/c
; Sequence 21331, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 21331
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT04380
US-10-750-185-21331

Query Match      66.4%; Score 16.6; DB 7; Length 600;
Best Local Similarity 82.6%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      3 AAAACTCCTGTAGGCTTTTCAGTG 25
Db      129 AATACTCCTGGGGCTTCCAGTG 107

RESULT 37
US-10-750-623-4713/c
; Sequence 4713, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
```

; PRIOR APPLICATION NUMBER: US 60/437,482  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 4713  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT04380  
US-10-750-623-4713

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 88;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25  
Db 129 AATACTCCTGGGGCTTCAGTG 107

RESULT 38  
US-10-750-623-21331/c  
; Sequence 21331, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 21331  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT04380  
US-10-750-623-21331

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 88;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25  
Db 129 AATACTCCTGGGGCTTCAGTG 107

RESULT 39  
US-10-750-185-26142  
; Sequence 26142, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 26142  
; LENGTH: 781  
; TYPE: DNA  
; ORGANISM: Bovine 19866881217231  
US-10-750-185-26142

Query Match 66.4%; Score 16.6; DB 7; Length 781;  
Best Local Similarity 82.6%; Pred. No. 93;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAATCCTGTAGGCTTTTCAGT 24  
Db 621 AAAAATCCTTGAAGGGTATCAGT 643

RESULT 40  
US-10-750-623-26142  
; Sequence 26142, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 26142  
; LENGTH: 781  
; TYPE: DNA  
; ORGANISM: Bovine 19866881217231  
US-10-750-623-26142

Query Match 66.4%; Score 16.6; DB 7; Length 781;  
Best Local Similarity 82.6%; Pred. No. 93;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAATCCTGTAGGCTTTTCAGT 24  
Db 621 AAAAATCCTTGAAGGGTATCAGT 643

RESULT 41  
US-10-750-185-52730/c  
; Sequence 52730, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 52730  
; LENGTH: 914

```

; TYPE: DNA
; ORGANISM: Bovine 19866881010885
; US-10-750-185-52730

Query Match      66.4%; Score 16.6; DB 7; Length 914;
Best Local Similarity 82.6%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAG 23
   ||||| ||||| ||||| |||||
Db 473 AACAACTCAGAGGCTTTTCAG 451

RESULT 42
US-10-750-623-52730/c
; Sequence 52730, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52730
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Bovine 19866881010885
US-10-750-623-52730

Query Match      66.4%; Score 16.6; DB 7; Length 914;
Best Local Similarity 82.6%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCCTGTAGGCTTTTCAG 23
   ||||| ||||| ||||| |||||
Db 473 AACAACTCAGAGGCTTTTCAG 451

RESULT 43
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-185-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 44
US-10-750-623-57290/c
; Sequence 57290, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-623-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 45
US-10-750-185-44378/c
; Sequence 44378, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 44378
; LENGTH: 1496
; TYPE: DNA
; ORGANISM: Bovine 19866880770781
US-10-750-185-44378

Query Match      66.4%; Score 16.6; DB 7; Length 1496;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 46
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-185-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 47
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-185-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 48
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-185-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 49
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 57290
; LENGTH: 1153
; TYPE: DNA
; ORGANISM: Bovine 19866881395807
US-10-750-185-57290

Query Match      66.4%; Score 16.6; DB 7; Length 1153;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAATCCTGTAGGCTTTTCAGTG 25
   ||||| ||||| ||||| |||||
Db 267 AATACTCTCGGGGCTTCCAGTG 245

RESULT 50
US-10-750-185-57290/c
; Sequence 57290, Application US/10750185
; Publication No. US20050260603A
```



Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAG 23  
|||||  
Db 303 AAAAAATCCTGTTGACTTGCG 281

## RESULT 46

US-10-750-623-44378/c  
; Sequence 44378, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 44378  
; LENGTH: 1496  
; TYPE: DNA  
; ORGANISM: Bovine 1986688070781  
US-10-750-623-44378

Query Match 66.4%; Score 16.6; DB 7; Length 1496;  
Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTCTGTAGGCTTTTCAG 23  
|||||  
Db 303 AAAAAATCCTGTTGACTTGCG 281

## RESULT 47

US-10-750-185-52605/c  
; Sequence 52605, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 52605  
; LENGTH: 1507  
; TYPE: DNA  
; ORGANISM: Bovine 19866880525142  
US-10-750-185-52605

Query Match 66.4%; Score 16.6; DB 7; Length 1507;  
Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTCTGTAGGCTTTTCAGT 24

Db 303 AAAAAATCCTTTTGGCAATTCAGT 281  
|||||

## RESULT 48

US-10-750-623-52605/c  
; Sequence 52605, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 52605  
; LENGTH: 1507  
; TYPE: DNA  
; ORGANISM: Bovine 19866880525142  
US-10-750-623-52605

Query Match 66.4%; Score 16.6; DB 7; Length 1507;  
Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTCTGTAGGCTTTTCAGT 24

Db 303 AAAAAATCCTTTTGGCAATTCAGT 281  
|||||

## RESULT 49

US-10-750-185-50030/c  
; Sequence 50030, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 50030  
; LENGTH: 1695  
; TYPE: DNA  
; ORGANISM: Bovine 19866880411509  
US-10-750-185-50030

Query Match 66.4%; Score 16.6; DB 7; Length 1695;  
Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTCTGTAGGCTTTTCAGT 24

Db 215 AAAAACTCTGAACGCTTTTCACT 193  
|||||

```
RESULT 50
US-10-750-623-50030/c
; Sequence 50030, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KEER, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50030
; LENGTH: 1695
; TYPE: DNA
; ORGANISM: Bovine 1986680411509
US-10-750-623-50030

Query Match      66.4%; Score 16.6; DB 7; Length 1695;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2  AAAAAGCTCTGTAGGCTTTCAGT 24
Db      215 AAAAAGCTCTGAACGCTTTCAC 193

Search completed: February 3, 2006, 16:20:07
Job time : 346.111 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-25  
Perfect score: 25  
Sequence: 1 aaaaactgcattcccaatgttatg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_in:.\*  
3: gb\_env:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pr:.\*  
9: gb\_ro:.\*  
10: gb\_sts:.\*  
11: gb\_sy:.\*  
12: gb\_un:.\*  
13: gb\_vi:.\*  
14: gb\_htg:.\*  
15: gb\_pl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.4	93.6	170951	9 AC123552	AC123552 Mus muscu
2	23.4	93.6	228147	9 AC090652	AC090652 Mus muscu
3	22.4	89.6	3146	9 RRP031145	XG2840 R.rattus MR
4	22.4	89.6	3977	6 AX839774	AX839774 Sequence
5	22.4	89.6	3977	6 AX839887	AX839887 Sequence
6	22.4	89.6	3977	9 RATKV4	MG8880 Rat neurona
C 7	22.4	89.6	217732	14 AC103001	AC103001 Rattus no
8	22.4	89.6	348540	14 AC132503	AC132503 Rattus no
9	21.8	87.2	270	6 AR270447	AR270447 Sequence
10	21.8	87.2	460	6 BD118543	BD118543 EST and e
11	21.8	87.2	460	6 AR422990	AR422990 Sequence
12	21.8	87.2	460	6 AX983684	AX983684 Sequence
13	21.8	87.2	1859	8 AX123157	AX123157 Homo sapi
C 14	21.8	87.2	115958	8 AC004736	AC004736 Human Chr
15	21.8	87.2	115971	8 AC124056	AC124056 Homo sapi
16	21.8	87.2	177142	8 AC004582	AC004582 Human Chr
17	21.8	87.2	177543	8 AC124078	AC124078 Homo sapi
18	20.8	83.2	109919	14 AP007327	AP007327 Lotus cor

19	20.4	81.6	1917	6 AR484315	AR484315 Sequence
20	20.4	81.6	1917	6 AX142745	AX142745 Sequence
21	20.4	81.6	3276	1 AF270200	AF270200 Staphyloc
22	20.4	81.6	3276	6 AR486154	AR486154 Sequence
23	20.4	81.6	3276	6 AX145518	AX145518 Sequence
C 24	20.4	81.6	110000	1 CP000029_20	Continuation (21 o
C 25	20.4	81.6	300698	1 AE016750	AE016750 Staphyloc
26	20.2	80.8	1382	5 BC099258	BC099258 Xenopus l
27	20.2	80.8	1731	5 BC077190	BC077190 Xenopus l
28	19.8	79.2	66122	8 AL732290	AL732290 Human DNA
29	19.8	79.2	148468	14 AL355882	AL355882 Homo sapi
C 30	19.8	79.2	148497	14 AC079592	AC079592 Homo sapi
31	19.8	79.2	185810	14 AC136052	AC136052 Rattus no
C 32	19.8	79.2	189389	8 CNS05TCF	AL355096 Human chr
33	19.8	79.2	211635	14 AC112763	AC112763 Rattus no
C 34	19.4	77.6	2124	5 BC045302	BC045302 Danio rer
C 35	19.4	77.6	46296	5 BX276107	BC276107 Zebrafish
36	19.4	77.6	58289	5 BX470142	BX470142 Zebrafish
37	19.4	77.6	140962	14 CR387991	CR387991 Danio rer
38	19.4	77.6	172243	9 AL606513	AL606513 Mouse DNA
C 39	19.4	77.6	185950	14 AL606486	AL606486 Homo sapi
C 40	19.4	77.6	228881	14 CR936412	CR936412 Danio rer
41	19.2	76.8	479	2 AY598742	AY598742 Trilobodr
C 42	19.2	76.8	694	10 BV660808	BV660808 S216P6047
C 43	19.2	76.8	959	10 BV524469	BV524469 G591P6048
C 44	19.2	76.8	988	10 BV543020	BV543020 sp143b08
C 45	19.2	76.8	3079	2 AB122062	AB122062 Crassosotr
C 46	19.2	76.8	39499	2 U41015	U41015 Caenorhabdi
C 47	19.2	76.8	110000	15 AP008209_256	Continuation (257
C 48	19.2	76.8	155293	14 AC154085	AC154085 Rhinoloph
49	19.2	76.8	157381	8 AL360267	AL360267 Human DNA
50	19.2	76.8	163328	14 AC016190	AC016190 Homo sapi
51	19.2	76.8	164394	9 AC103388	AC103388 Mus muscu
C 52	19.2	76.8	164690	14 AC016846	AC016846 Homo sapi
C 53	19.2	76.8	166437	14 AC135133	AC135133 Rattus no
54	19.2	76.8	167427	15 AC139174	AC139174 Oryza sat
55	19.2	76.8	170756	5 BX248508	BX248508 Zebrafish
C 56	19.2	76.8	171245	14 CR762414	CR762414 Danio rer
57	19.2	76.8	173432	14 AC141166	AC141166 Rattus no
C 58	19.2	76.8	179144	5 BX511025	BX511025 Zebrafish
59	19.2	76.8	179651	8 AC005553	AC005553 Homo sapi
60	19.2	76.8	187298	8 AC018691	AC018691 Homo sapi
61	19.2	76.8	188004	14 AC113894	AC113894 Rattus no
62	19.2	76.8	189055	8 AC015724	AC015724 Homo sapi
63	19.2	76.8	198903	8 BS000070	BS000070 Pan trogl
64	19.2	76.8	200521	9 AL833804	AL833804 Mouse DNA
C 65	19.2	76.8	203959	8 AP001137	AP001137 Homo sapi
C 66	19.2	76.8	206136	8 AC068037	AC068037 Homo sapi
67	19.2	76.8	219117	9 AC129332	AC129332 Mus muscu
68	19.2	76.8	222594	14 AC124931	AC124931 Rattus no
69	19.2	76.8	224716	14 AC133852	AC133852 Rattus no
C 70	19.2	76.8	225959	14 CR925879	CR925879 Danio rer
C 71	19.2	76.8	232537	14 AC157781	AC157781 Mus muscu
C 72	19.2	76.8	238807	14 AC116254	AC116254 Rattus no
73	19.2	76.8	239423	14 AC096022	AC096022 Rattus no
C 74	19.2	76.8	248921	14 AC095538	AC095538 Rattus no
75	19.2	76.8	268898	14 AC109095	AC109095 Rattus no
76	19.2	76.8	340000	8 AP001681	AP001681 Homo sapi
77	19	76.0	159468	5 AL935040	AL935040 Zebrafish
78	19	76.0	170054	14 BX537266	BX537266 Danio rer
79	19	76.0	173631	5 BX537275	BX537275 Zebrafish
80	18.8	75.2	307	10 CFA411044	CFA411044 Canis fam
C 81	18.8	75.2	589	10 BV429969	BV429969 S237P6264
C 82	18.8	75.2	717	10 BV517218	BV517218 qdv77c02
C 83	18.8	75.2	729	5 CR406939	CR406939 Gallus ga
C 84	18.8	75.2	771	10 BV535492	BV535492 G591P6286
85	18.8	75.2	775	10 BV527407	BV527407 G591P6108
86	18.8	75.2	786	10 BV535377	BV535377 G591P6301
87	18.8	75.2	815	10 BV517667	BV517667 rln39c05
C 88	18.8	75.2	856	10 BV512020	BV512020 qbn32d10
89	18.8	75.2	53542	6 AX247730	AX247730 Sequence
90	18.8	75.2	96387	14 HSJ600024	AL080247 Homo sapi
C 91	18.8	75.2	109374	8 AP000952	AP000952 Homo sapi

92	18.8	75.2	110000	15	AP008217_181	Continuation (182
93	18.8	75.2	110000	15	AP008212_134	Continuation (135
94	18.8	75.2	136799	15	AP003506	AP003506 Oryza sat
95	18.8	75.2	143983	8	AC099526	AC099526 Papio anu
96	18.8	75.2	162413	15	AC124150	AC124150 Oryza sat
97	18.8	75.2	168230	15	AP003572	AP003572 Oryza sat
98	18.8	75.2	175377	14	AC120929	AC120929 Rattus no
99	18.8	75.2	175492	8	AC036101	AC036101 Homo sapi
100	18.8	75.2	175980	14	AC025082	AC025082 Homo sapi
101	18.8	75.2	177241	8	HS402G11	AL022328 Human DNA
102	18.8	75.2	185702	9	AC122817	AC122817 Mus muscu
103	18.8	75.2	196646	9	AC125174	AC125174 Mus muscu
104	18.8	75.2	206673	14	CR327368	CR327368 Danio rer
105	18.8	75.2	225055	8	BS000035	BS000035 Pan trogl
106	18.8	75.2	233095	14	AC094266	AC094266 Rattus no
107	18.8	75.2	267054	14	AC094354	AC094354 Rattus no
108	18.8	75.2	340000	8	AP001670	AP001670 Homo sapi
109	18.8	75.2	342537	14	AC096912	AC096912 Rattus no
110	18.6	74.4	549	10	G27795	G27795 human STS S
111	18.6	74.4	640	10	BV370765	BV370765 S231P6217
112	18.6	74.4	740	10	BV651065	BV651065 S217P6203
113	18.6	74.4	3613	13	TIDG1G2A	L08756 Thailand vi
114	18.6	74.4	3619	8	BC036207	BC036207 Homo sapi
115	18.6	74.4	4706	8	AY532641	AY532641 Pongo pyg
116	18.6	74.4	31682	2	U64601	U64601 Caenorhabdi
117	18.6	74.4	43241	4	AC155903	AC155903 Ixodonta
118	18.6	74.4	66152	14	AC090613	AC090613 Homo sapi
119	18.6	74.4	74500	1	AE009951_21	Continuation (22 o
120	18.6	74.4	74500	1	AE009951_21	Continuation (22 o
121	18.6	74.4	81386	8	AC121320	AC121320 Homo sapi
122	18.6	74.4	84619	14	AC009942	AC009942 Homo sapi
123	18.6	74.4	89275	14	AP007683	AP007683 lotus cor
124	18.6	74.4	91254	14	AP007856	AP007856 lotus cor
125	18.6	74.4	102914	14	AP007872	AP007872 lotus cor
126	18.6	74.4	110000	1	AE009951_00	AE009951 Fusobacte
127	18.6	74.4	110000	1	AE009951_01	Continuation (2 of
128	18.6	74.4	110000	1	AE009951_03	Continuation (4 of
129	18.6	74.4	110000	1	AE009951_04	Continuation (5 of
130	18.6	74.4	110000	1	AE009951_07	Continuation (8 of
131	18.6	74.4	110000	1	AE009951_11	Continuation (12 o
132	18.6	74.4	110000	1	AE009951_14	Continuation (15 o
133	18.6	74.4	110000	1	AE009951_15	Continuation (16 o
134	18.6	74.4	110000	15	AP008217_260	Continuation (261
135	18.6	74.4	130497	14	AC165215	AC165215 Alligator
136	18.6	74.4	131946	15	AC129226	AC129226 Oryza sat
137	18.6	74.4	132938	9	AY900632	AY900632 Drosophil
138	18.6	74.4	135978	9	AL645582	AL645582 Mouse DNA
139	18.6	74.4	154354	8	AC105923	AC105923 Homo sapi
140	18.6	74.4	156980	14	AC068032	AC068032 Homo sapi
141	18.6	74.4	157228	5	CR354536	CR354536 Zebraphis
142	18.6	74.4	158930	14	AC022531	AC022531 Homo sapi
143	18.6	74.4	159939	5	BX649532	BX649532 Zebraphis
144	18.6	74.4	160654	14	AC011879	AC011879 Homo sapi
145	18.6	74.4	162391	8	AC090577	AC090577 Homo sapi
146	18.6	74.4	163152	8	CNS05TDI	AL356800 Human chr
147	18.6	74.4	166133	8	AC092361	AC092361 Homo sapi
148	18.6	74.4	167133	14	AC026186	AC026186 Homo sapi
149	18.6	74.4	168174	6	AR584462	AR584462 Sequence
150	18.6	74.4	168239	9	AC151899	AC151899 Mus muscu

## ALIGNMENTS

RESULT 1	AC123552	170951 bp	DNA	linear	ROD 21-NOV-2003
LOCUS	Mus musculus chromosome 7 clone RP24-396118				complete sequence.
DEFINITION	AC123552				
ACCESSION	AC123552				
VERSION	AC123552.4	GI:38488886			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				

REFERENCE	1 (bases 1 to 170951)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	Wilson,R.K.	
TITLE	The sequence of Mus musculus clone	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 170951)	
AUTHORS	McPherson,J.D. and Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
REFERENCE	3 (bases 1 to 170951)	
AUTHORS	McPherson,J.D. and Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-MAR-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
REFERENCE	4 (bases 1 to 170951)	
AUTHORS	Wilson,R.K.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
COMMENT	On Nov 21, 2003 this sequence version replaced gi:28828036.	
FEATURES	Location/Qualifiers	
source	1..170951	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10090"	
	/chromosome="7"	
	/clone="RP24-396118"	
ORIGIN		
Query Match	93.6%; Score 23.4; DB 9; Length 170951;	
Best Local Similarity	96.0%; Pred. No. 3;	
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AAAAACTGCATCCCAATGTTATG 25	
Db	157672 AAAAACTGCATCCCAATGTTATG 157648	
RESULT 2	AC090652	228147 bp
LOCUS	Mus musculus strain C57BL/6J chromosome 7 clone rp23-116a10,	DNA linear ROD 27-NOV-2002
DEFINITION	complete sequence.	
ACCESSION	AC090652	
VERSION	AC090652.32	GI:25703442
KEYWORDS	HTG.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	1 (bases 1 to 228147)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
AUTHORS	Jiang,X., Song,L. and Roe,B.A.	
TITLE	Mus musculus BAC clone rp23-116a10	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 228147)	
AUTHORS	Jiang,X., Song,L. and Roe,B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (07-MAR-2001) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	

REFERENCE	3 (bases 1 to 228147)	
AUTHORS	Jiang, X., Song, L. and Roe, B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-NOV-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	
REFERENCE	4 (bases 1 to 228147)	
AUTHORS	Jiang, X., Song, L. and Roe, B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-NOV-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	
REFERENCE	5 (bases 1 to 228147)	
AUTHORS	Jiang, X., Song, L. and Roe, B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (27-NOV-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	
COMMENT	On Nov 27, 2002 this sequence version replaced gi:24961194. ----- Genome Center Center: Department Of Chemistry And Biochemistry The University Of Oklahoma Center code:UOKNOR -----	
FEATURES	Location/Qualifiers	
source	1..228147	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/chromosome="7"	
	/clone="rp23-116a10"	
	/clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"	
ORIGIN		
Query Match	93.6%; Score 23.4; DB 9; Length 228147;	
Best Local Similarity	96.0%; Pred. No. 3;	
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AAAAACTGCATCCCAATGTTATG 25	
Db	29835 AAAAACTGCATCCCAATGTTATG 29859	
RESULT 3		
RRPCP3145		
LOCUS	3146 bp mRNA linear ROD 18-APR-2005	
DEFINITION	R.rattus mRNA for potassium channel protein (3145 bp).	
ACCESSION	X62840	
VERSION	X62840.1 GI:57652	
KEYWORDS	voltage-gated potassium channel.	
SOURCE	Rattus rattus (black rat)	
ORGANISM	Rattus rattus	
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE	1 (bases 1 to 3146)	
AUTHORS	Rattling, J., Wunder, F., Stocker, M., Lichtinghagen, R., Mastiaux, F., Beckh, S., Kues, W., Pedarzani, P., Schroeter, K.H., Ruppersberg, J.P., Veh, R. and Pongs, O.	
TITLE	Characterization of a Shaw-related potassium channel family in rat brain	
JOURNAL	EMBO J. 11 (7), 2473-2486 (1992)	
PUBMED	1378392	
REFERENCE	2 (bases 1 to 3146)	
AUTHORS	Pongs, O.	
TITLE	Direct Submission	
JOURNAL	Submitted (22-OCT-1991) O. Pongs, Zentrum f Molekulare Neurobiologie, ZNKH Haus 42, Martinistr 52, 2000 Hamburg 20, FRG	
FEATURES	Location/Qualifiers	
source	1..3146	
	/organism="Rattus rattus"	
	/mol_type="mRNA"	
/db_xref="taxon:10117"		
/clone_lib="cortex cDNA, gt10"		
1..3103		
/evidence=experimental		
287..2044		
/codon_start=1		
/product="voltage-gated potassium channel"		
/protein_id="CAA44644.1"		
/db_xref="GI:57653"		
/db_xref="GOA:P25122"		
/db_xref="InterPro:IPR001622"		
/db_xref="InterPro:IPR003091"		
/db_xref="InterPro:IPR003131"		
/db_xref="InterPro:IPR003968"		
/db_xref="InterPro:IPR003974"		
/db_xref="InterPro:IPR005403"		
/db_xref="InterPro:IPR005820"		
/db_xref="InterPro:IPR005821"		
/db_xref="UniProt/Swiss-Prot:P25122"		
/translation="MQGDESERIVINVGTRHQTYSRTRTLPGTRLAWLPDAHS HFYDPRADAEFFDRHPGVFAHILNYRTKLCADVCGLYEEELAFWIDTDE PCWMTYRQHRDAEALDSFGGAPLDSADADADGDSGDGEDELEMTKRLALSDS PDGRPGFWRWQPRIWALFEDPYSSRYARYVAFSLFFILVITTFCLTHERFNPI VNKTEIENVNGTQVRYREARETEAFLEIGVCVVFTEFLMRVVCPCNKVEFLKN SLNIIDFVAILPFYLVGSLSSKAQVGLFRLVRVRLIRIFKLRHFFVGLRVL GHTLRASNFELLLIFLAGLVLIFATMIYAAERIGAGNDPSASHTHFPKNIPIGW WAVVTMTLGYGDMYPQTMGLMVGALCALAGVLTIAMPPVIVNNFVYSLAMAKQ KLPKKKKHIPRPOLGSPNYCKSVNSPHHSQSDTCLPAOEIILEINRADSKLNGE VAKAALANEDCPHIDQALTPDEGLPTRSGTRERYPCPLFLLSTGEVACPPGGWKRDL CKESPVIKMYMTEAVRVV"		
3103		
polyA_site		
ORIGIN		
Query Match	89.6%; Score 22.4; DB 9; Length 3146;	
Best Local Similarity	95.8%; Pred. No. 12;	
Matches	23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	2 AAAAACTGCATCCCAATGTTATG 25	
Db	3042 AAAAACTGCATCCCAATGTTATG 3065	
RESULT 4		
AX839774		
LOCUS	3977 bp DNA linear PAT 16-DEC-2003	
DEFINITION	Sequence 1 from Patent EP1348963.	
ACCESSION	AX839774	
VERSION	AX839774.1 GI:39978333	
KEYWORDS	Rattus norvegicus (Norway rat)	
SOURCE	Rattus norvegicus	
ORGANISM	Rattus norvegicus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE	1	
AUTHORS	Cochran, S.M., Yamagami, K.M. and Ohashi, Y.M.	
TITLE	Schizophrenia related gene	
JOURNAL	Patent: EP 1348963-A 1 01-OCT-2003;	
	Mitsubishi Pharma Corporation (JP)	
FEATURES	Location/Qualifiers	
source	1..3977	
	/organism="Rattus norvegicus"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:10116"	
	1162..2919	
	/notes="unnamed protein product"	
	/codon_start=1	
	/protein_id="CAB85419.1"	
	/db_xref="GI:39978334"	
	/translation="MQGDESERIVINVGTRHQTYSRTRTLPGTRLAWLPDAHS HFYDPRADAEFFDRHPGVFAHILNYRTKLCADVCGLYEEELAFWIDTDE PCWMTYRQHRDAEALDSFGGAPLDSADADADGDSGDGEDELEMTKRLALSDS PDGRPGFWRWQPRIWALFEDPYSSRYARYVAFSLFFILVITTFCLTHERFNPI	

```

VNKTEIENVRNGTQVRYREAEATEAFLTYIEGVCVVWFTEBFLMRVVFPCNKKVEFIKN
SLNIIDFVAIILPFYLEVGLSGSSKAADVLGFLRVRFVRLIRIFKLTIRHFVGLRVL
GHTLRASTEFLLLIIFALGVLIIFATMIYYAERIGAQPNPDSASEHFKNIPIGFV
WAVVTMTTIGYDMYPTQWSGMLVGCALAGVLTAMPVPVIVNNFGMYSLAMAKQ
KLPKKKKKHIPRPOLGSPNYCKSVNSPHSTQSDTCPLAQEETLEINRADSKLNGE
VAKAALANEDCPHIDQALTPDEGLPFTSRGRYRGCPCFLLSTGEGYACPPGGMRKDL
CKESPVIAYKMPTEAVRVVT"
3958. .3963

polyA_signal
ORIGIN
Query Match      89.6%; Score 22.4; DB 6; Length 3977;
Best Local Similarity 95.8%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 AAAAACTGCATCCCAATGTTATG 25
|||||
Db  3915 AAAAACTGCATCCCAATGTTATG 3938

RESULT 5
AX839887
LOCUS      AX839887          3977 bp      DNA      linear      PAT 16-DEC-2003
DEFINITION Sequence 1 from Patent EP1348964.
ACCESSION  AX839887
VERSION     AX839887.1  GI:39978418
KEYWORDS
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Rattus.
1
REFERENCE   1
AUTHORS    Cochran,S.Y., Yamagami,K.M. and Ohashi,Y.M.
TITLE       Schizophrenia related gene
JOURNAL     Patent: EP 1348964-A 1 01-OCT-2003;
            Mitsubishi Pharma Corporation (JP)
FEATURES    Location/Qualifiers
            source
            1..3977
            /organism="Rattus norvegicus"
            /mol_type="unassigned DNA"
            /db_xref="taxon:10116"
            1162..2919
            /note="unamed protein product"
            /codon_start=1
            /protein_id="CAE85433.1"
            /db_xref="GI:39978419"
            1162..2919
            /translat="MGQGDSESRIVNVTGTHQYRSTLRTLPGLRLAWLAEPDAHS
            HFDYPRADRFPPDRHGPVFAHILNRYRTGKLHCADVPGLYEEELAFWGIDETIVE
            PCWMTYRQHRDAEALDSFGGAPLONSADADAGPGSGDGELEMTKRLALSDS
            PDGREGFWRRQPRIWALFEDPYSRARYVAFSLFLLVSIITFCLETHERNPI
            VNKTEIENVRNGTQVRYREAEATEAFLTYIEGVCVVWFTEBFLMRVVFPCNKKVEFIKN
            SLNIIDFVAIILPFYLEVGLSGSSKAADVLGFLRVRFVRLIRIFKLTIRHFVGLRVL
            GHTLRASTEFLLLIIFALGVLIIFATMIYYAERIGAQPNPDSASEHFKNIPIGFV
            WAVVTMTTIGYDMYPTQWSGMLVGCALAGVLTAMPVPVIVNNFGMYSLAMAKQ
            KLPKKKKKHIPRPOLGSPNYCKSVNSPHSTQSDTCPLAQEETLEINRADSKLNGE
            VAKAALANEDCPHIDQALTPDEGLPFTSRGRYRGCPCFLLSTGEGYACPPGGMRKDL
            CKESPVIAYKMPTEAVRVVT"
            3958. .3963

polyA_signal
ORIGIN
Query Match      89.6%; Score 22.4; DB 6; Length 3977;
Best Local Similarity 95.8%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 AAAAACTGCATCCCAATGTTATG 25
|||||
Db  3915 AAAAACTGCATCCCAATGTTATG 3938

RESULT 6
RATKV4
LOCUS      RATKV4          3977 bp      mRNA      linear      ROD 27-APR-1993
DEFINITION Rat neuronal delayed rectifier K+ channel (K-V-4) mRNA, complete
```

```

cds.
M68880 M37845
M68880.1 GI:205106
neural delayed rectifier K+ channel.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 3977)
Luneau,C.J., Williams,J.B., Marshall,J., Levitan,E.S., Oliva,C.,
Smith,J.S., Antanavage,J., Foltander,K., Stein,R.B., Swanson,R.,
Kaczmarek,L.K. and Buhrow,S.A.
Alternative splicing contributes to K+ channel diversity in the
mammalian central nervous system
Proc. Natl. Acad. Sci. U.S.A. 88 (9), 3932-3936 (1991)
2023941
COMMENT    Original source text: Rattus norvegicus (strain Sprague-Dawley)
            brain cDNA to mRNA.
FEATURES    Location/Qualifiers
            source
            1..3977
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /tissue_type="brain"
            1..3977
            /gene="K-V-4"
            /gene="K-V-4"
            1..1161
            /gene="K-V-4"
            1162..2919
            /gene="K-V-4"
            /codon_start=1
            /product="neuronal delayed rectifier K+ channel"
            /protein_id="AAA41501.1"
            /db_xref="GI:205107"
            /translat="MGQGDSESRIVNVTGTHQYRSTLRTLPGLRLAWLAEPDAHS
            HFDYPRADRFPPDRHGPVFAHILNRYRTGKLHCADVPGLYEEELAFWGIDETIVE
            PCWMTYRQHRDAEALDSFGGAPLONSADADAGPGSGDGELEMTKRLALSDS
            PDGREGFWRRQPRIWALFEDPYSRARYVAFSLFLLVSIITFCLETHERNPI
            VNKTEIENVRNGTQVRYREAEATEAFLTYIEGVCVVWFTEBFLMRVVFPCNKKVEFIKN
            SLNIIDFVAIILPFYLEVGLSGSSKAADVLGFLRVRFVRLIRIFKLTIRHFVGLRVL
            GHTLRASTEFLLLIIFALGVLIIFATMIYYAERIGAQPNPDSASEHFKNIPIGFV
            WAVVTMTTIGYDMYPTQWSGMLVGCALAGVLTAMPVPVIVNNFGMYSLAMAKQ
            KLPKKKKKHIPRPOLGSPNYCKSVNSPHSTQSDTCPLAQEETLEINRADSKLNGE
            VAKAALANEDCPHIDQALTPDEGLPFTSRGRYRGCPCFLLSTGEGYACPPGGMRKDL
            CKESPVIAYKMPTEAVRVVT"
            2920..3977
            /gene="K-V-4"
            3958..3963
            /gene="K-V-4"

polyA_signal
ORIGIN
Query Match      89.6%; Score 22.4; DB 9; Length 3977;
Best Local Similarity 95.8%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 AAAAACTGCATCCCAATGTTATG 25
|||||
Db  3915 AAAAACTGCATCCCAATGTTATG 3938

RESULT 7
AC103001/c
LOCUS      AC103001          21732 bp      DNA      linear      HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-185F19, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION  AC103001
VERSION     AC103001.8  GI:30581312
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Rattus.

# REFERENCE AUTHORS

1 (bases 1 to 217732)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, R., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gragegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, H., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Monmayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindestre, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

# TITLE

# REFERENCE AUTHORS

2 (bases 1 to 217732)  
Worley, K. C.  
Direct Submission  
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 217732)

# REFERENCE AUTHORS

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:24942563.

# COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GIRU  
Center clone name: CH230-185F19  
----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 184743 bases at least Q40  
Consensus quality: 185828 bases at least Q30  
Consensus quality: 186720 bases at least Q20  
Estimated insert size: 189327; sum-of-contigs estimation  
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 215333: contig of 215333 bp in length  
\* 215334 215433: gap of unknown length  
\* 215334 216477: contig of 1044 bp in length  
\* 216478 216577: gap of unknown length  
\* 216578 217732: contig of 1155 bp in length.

# FEATURES

## source

1. 217732  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-185F19"

## misc\_feature

1. 1505  
/notes="wgs end extension  
clone\_end:T7"  
2095. 4056  
/notes="wgs end extension  
clone\_end:T7"

## misc\_feature

complement(5859. 6364)  
/notes="clone boundary  
clone\_end:T7  
site:ECORI  
end sequence: BH263208"  
complement(213207. 213763)  
/notes="clone boundary  
clone\_end:Sp6  
site:ECORI  
end sequence: BH263257"

## misc\_feature

215334. 215433  
/estimated\_length=unknown  
216478. 216577  
/estimated\_length=unknown

## misc\_feature

215334. 215433  
/estimated\_length=unknown  
216478. 216577  
/estimated\_length=unknown

## gap

## gap

# ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 217732;  
Best Local Similarity 95.8%; Pred. No. 9;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

# Qy

2 AAAAAGTCATCCCAATGTTATG 25  
|||||

# Db

93910 AAAAAGTCATCCCAATGTTATG 93887

Mon Feb 6 12:23:14 2006

AUTHORS  
TITLE  
JOURNAL

Rat Genome Sequencing Consortium.

## COMMENT

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23908470. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KBRQ

Center clone name: CH230-425P24

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 295693 bases at least Q40

Consensus quality: 301215 bases at least Q30

Consensus quality: 304883 bases at least Q20

Estimated insert size: 295566; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases, sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)

\* NOTE: This sequence may represent more than one clone.

\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 31755: contig of 31755 bp in length

31756: gap of unknown length

31855: contig of 29070 bp in length

60925: gap of unknown length

60926: gap of unknown length

61025: gap of unknown length

61026: gap of unknown length

64397: gap of unknown length

64398: contig of 65823 bp in length

130320: gap of unknown length

130321: contig of 16156 bp in length

130421: gap of unknown length

146577: gap of unknown length

146577: gap of unknown length

157802: contig of 11126 bp in length

157902: gap of unknown length

157903: contig of 188180 bp in length

346082: gap of unknown length

346182: gap of unknown length

347276: contig of 1094 bp in length

347277: gap of unknown length

348540: contig of 1164 bp in length.

347377 Location/Qualifiers

1. 348540

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-425P24"

11273..11359

/note="clone\_boundary"

clone\_end:Sp6

site:

end\_sequence:BZ211011"

FEATURES  
source

1. 348540

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-425P24"

11273..11359

/note="clone\_boundary"

clone\_end:Sp6

site:

end\_sequence:BZ211011"

## misc\_feature

11273..11359

/note="clone\_boundary"

clone\_end:Sp6

site:

end\_sequence:BZ211011"

TITLE  
JOURNAL

Rat Genome Sequencing Consortium.

AUTHORS  
TITLE  
JOURNAL

Submitted (01-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

2 (bases 1 to 348540)

Rat Genome Sequencing Consortium.

Direct Submission

## REFERENCE

Submitted (01-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

2 (bases 1 to 348540)

## RESULT 8

AC132503

LOCUS

## DEFINITION

Rattus norvegicus clone CH230-425P24, \*\*\* SEQUENCING IN PROGRESS

## ACCESSION

AC132503.3 GI:25139153

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

## KEYWORDS

Rattus norvegicus (Norway rat)

## SOURCE

Rattus norvegicus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus;

## REFERENCE

1 (bases 1 to 348540)

## AUTHORS

Muzny, D. Marie., Metaker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, N., Hernandez, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, C., Hamilton, K., Hollings, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuher, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quatroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, F., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.

```

misc_feature 23469..24939
              /note="wgs contig"
misc_feature 24990..26503
              /note="wgs contig"
gap 31756..31855
      /estimated_length=unknown
misc_feature 31856..32991
              /note="wgs contig"
misc_feature 33343..34933
              /note="wgs contig"
gap 60926..61025
      /estimated_length=unknown
misc_feature 61026..62237
              /note="wgs contig"
gap 64398..64497
      /estimated_length=unknown
misc_feature 102826..106931
              /note="wgs contig"
gap 130321..130420
      /estimated_length=unknown
misc_feature 130421..131444
              /note="wgs contig"
gap 146577..146676
      /estimated_length=unknown
misc_feature 156232..157802
              /note="wgs contig"
gap 157803..157902
      /estimated_length=unknown
misc_feature 157903..158937
              /note="wgs contig"
misc_feature 161582..163160
              /note="wgs contig"
misc_feature 184173..184811
              /note="clone_boundary"
              /note="clone_site"
              /end_sequence=BZ211009"
misc_feature 343981..346082
              /note="wgs_end_extension"
              /clone_end=T7"
gap 346083..346182
      /estimated_length=unknown
gap 347277..347376
      /estimated_length=unknown

ORIGIN
Query Match 89.6%; Score 22.4; DB 14; Length 348540;
Best Local Similarity 95.8%; Pred. No. 8.7;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
    |||||
DB 237083 AAAAACTGCATCCCAATGTTATG 237106

RESULT 9
AR270447 LOCUS 270 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 1010 from patent US 6500938.
ACCESSION AR270447
VERSION AR270447.1 GI:29701681
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 270)
AUTHORS Au-Young, J. and Seilhamer, J.J.
TITLE Composition for the detection of signaling pathway gene expression
JOURNAL Patent: US 6500938-A 1010 31-DEC-2002;
Incyte Genomics, Inc.; Palo Alto, CA;
WOX;
FEATURES
    _source Location/Qualifiers
        1..270

/organism="unknown"
/mol_type="genomic DNA"

Query Match 87.2%; Score 21.8; DB 6; Length 270;
Best Local Similarity 92.0%; Pred. No. 28;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    |||||
DB 164 AAAAACTGCATCCCAATGTTATG 188

RESULT 10
BD118543 LOCUS 460 bp DNA linear PAT 18-SEP-2002
DEFINITION EST and encoded human protein.
ACCESSION BD118543
VERSION BD118543.1 GI:23213453
KEYWORDS JP 2002010789-A/10620.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Hominidae; Homo.
REFERENCE 1 (bases 1 to 460)
AUTHORS Edwards, J.B.D.M., Jobert, S. and Giordano, J.E.
TITLE EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 10620 15-JAN-2002;
GENSET CORP
COMMENT OS Homo sapiens (human)
PN JP 2002010789-A/10620
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 05-AUG-1999 US 60/147499
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FH Key Location/Qualifiers
FT source 1..460
FT /organism="Homo sapiens (human)".

FEATURES
    source Location/Qualifiers
        1..460
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"

ORIGIN
Query Match 87.2%; Score 21.8; DB 6; Length 460;
Best Local Similarity 92.0%; Pred. No. 27;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    |||||
DB 136 AAAAACTGCATCCCAATGTTATG 160

RESULT 11
AR422990 LOCUS 460 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 14487 from patent US 6639063.
ACCESSION AR422990
VERSION AR422990.1 GI:40178100
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 460)
AUTHORS Edwards, J.-B.D.M., Jobert, S. and Giordano, J.-Y.

```

```

TITLE      EST's and encoded human proteins
JOURNAL    Patent: US 6639063-A 14487 28-OCT-2003;
           Genset S.A.;;
FEATURES   Location/Qualifiers
source     1. .460
           /organism="unknown"
           /mol_type="genomic DNA"
ORIGIN
Query Match      87.2%; Score 21.8; DB 6; Length 460;
Best Local Similarity 92.0%; Pred. No. 27;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    ||||| ||||| ||||| ||||| |||||
Db 136 AAAAACTGCATCCCAATGTTATG 160

RESULT 12
AX983684      460 bp      DNA      PAT 15-JAN-2004
LOCUS         Sequence 14487 from Patent EP1104808.
ACCESSION     AX983684
VERSION       AX983684.1 GI:40989824
KEYWORDS      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominoidea; Homo.
REFERENCE     1
AUTHORS       Dumas Milne Edwards,J.B., Jobert,S. and Giordano,J.Y.
TITLE         ESTs and encoded human proteins
JOURNAL       Patent: EP 1104808-A 14487 06-JUN-2001;
              Genset (FR)
FEATURES      Location/Qualifiers
source       1. .460
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"
ORIGIN
Query Match      87.2%; Score 21.8; DB 6; Length 460;
Best Local Similarity 92.0%; Pred. No. 27;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    ||||| ||||| ||||| ||||| |||||
Db 136 AAAAACTGCATCCCAATGTTATG 160

RESULT 13
AK123157      1859 bp      mRNA      linear      PRI 30-JAN-2004
LOCUS         Homo sapiens cDNA FLJ41162 fis, clone BRACE2039600, highly similar
DEFINITION    to VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.1.
ACCESSION     AK123157
VERSION       AK123157.1 GI:34528632
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE        Homo sapiens
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominoidea; Homo.
REFERENCE     1
AUTHORS       Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
              Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
              Sekine,M., Oobayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
              Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
              Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
              Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
              Sugawara,M., Takahashi,M., Kanda,K., Kanda,K., Yokoi,T., Furuya,T.,

```

```

Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Kanehori,S., Komai,F., Hara,R.,
Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shichata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togaishi,T.,
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2
Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuoka,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 1859)
Isogai,T. and Yamamoto,Y.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES      Location/Qualifiers
source       1. .1859
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="BRACE2039600"
              /tissue_type="cerebellum"
              /clone_lib="BRACE2"
              /note="cloning vector: pME18SFL3"
ORIGIN
Query Match      87.2%; Score 21.8; DB 8; Length 1859;
Best Local Similarity 92.0%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    ||||| ||||| ||||| ||||| |||||
Db 1472 AAAAACTGCATCCCAATGTTATG 1496

RESULT 14
AC004736/c
LOCUS         AC004736      115958 bp      DNA      linear      PRI 03-JUL-1998

```

DEFINITION	Human Chromosome 11p14.3 PAC clone pDJ1082L12 containing KNCN1 and MyoD, complete sequence.	FEATURES	Location/Qualifiers
ACCESSION	AC004736	source	1. .115958
VERSION	AC004736.1 GI:3282183		/organism="Homo sapiens"
KEYWORDS	HTG.		/mol_type="genomic DNA"
SOURCE	Homo sapiens (human)		/db_xref="taxon:9606"
ORGANISM	Homo sapiens	repeat_region	4409. .4517
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.	repeat_region	/rpt_family="MIR"
REFERENCE	1 (bases 1 to 115958)	repeat_region	complement(19171. .19291)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Alu"
	HTGS Submission	repeat_region	23678. .23767
TITLE	Unpublished	repeat_region	/rpt_family="MIR"
REFERENCE	2 (bases 1 to 115958)	repeat_region	28877. .31250
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Tigger1"
	Direct Submission	repeat_region	complement(31777. .31876)
TITLE	Submitted (23-MAY-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	3 (bases 1 to 115958)	repeat_region	complement(36023. .36231)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	complement(39712. .39766)
TITLE	Submitted (01-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	4 (bases 1 to 115958)	repeat_region	complement(41971. .42073)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	45642. .45752
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	5 (bases 1 to 115958)	repeat_region	57204. .57337
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	complement(58027. .58249)
TITLE	Submitted (01-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MER20"
REFERENCE	6 (bases 1 to 115958)	repeat_region	59316. .59729
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="L1"
	Direct Submission	repeat_region	complement(60013. .60386)
TITLE	Submitted (01-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="THE1"
REFERENCE	7 (bases 1 to 115958)	repeat_region	60398. .60596
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="L1"
	Direct Submission	repeat_region	60940. .61228
TITLE	Submitted (01-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="Alu"
REFERENCE	8 (bases 1 to 115958)	repeat_region	complement(63635. .63900)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Alu"
	Direct Submission	repeat_region	complement(75510. .75774)
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="Alu"
REFERENCE	9 (bases 1 to 115958)	repeat_region	complement(82106. .82378)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Alu"
	Direct Submission	repeat_region	complement(84150. .84211)
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MLT1"
REFERENCE	10 (bases 1 to 115958)	repeat_region	84727. .85172
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MLT1"
	Direct Submission	repeat_region	85852. .86134
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="Alu"
REFERENCE	11 (bases 1 to 115958)	repeat_region	86387. .86565
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Alu"
	Direct Submission	repeat_region	complement(88437. .88514)
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	12 (bases 1 to 115958)	repeat_region	complement(90686. .90752)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	91042. .91163
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	13 (bases 1 to 115958)	repeat_region	94274. .94419
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	complement(95470. .95617)
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MLT1"
REFERENCE	14 (bases 1 to 115958)	repeat_region	complement(97484. .97917)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="L1"
	Direct Submission	repeat_region	105586. .105657
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="MIR"
REFERENCE	15 (bases 1 to 115958)	repeat_region	105673. .105884
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="MIR"
	Direct Submission	repeat_region	complement(106369. .106641)
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="Alu"
REFERENCE	16 (bases 1 to 115958)	repeat_region	complement(107302. .107596)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.	repeat_region	/rpt_family="Alu"
	Direct Submission	repeat_region	109188. .109458
TITLE	Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA	repeat_region	/rpt_family="Alu"

MAPPED CLONE OVERLAP: pDJ59m18 and 6-109h6.  
 IMPORTANT: This submission contains the entire insert of clone pDJ1082L12. pDJ1082L12 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected phred/phrap calculated errors/10kb is 0.39. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

```

repeat_region complement(110473. .110773)
/rpt_family="Alu"
repeat_region complement(111227. .111337)
/rpt_family="L1"
repeat_region 111472. .111792
/rpt_family="THE1"
repeat_region complement(112629. .112992)
/rpt_family="L1"

ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 115958;
Best Local Similarity 92.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
||||| ||||| ||||| ||||| |||||
Db 6193 AAAAACTGCATGCCCAATGTTATG 6169

RESULT 15
AC124056 115971 bp DNA linear PRI 21-DEC-2002
LOCUS Homo sapiens chromosome 11, clone RPS-1082L12, complete sequence.
DEFINITION AC124056
ACCESSION AC124056.8 GI:27356724
VERSION HTG
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eumetazoa; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.

REFERENCE
1 (bases 1 to 115971)
Birren,B., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RPS-1082L12
Unpublished
2 (bases 1 to 115971)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,T., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigliio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
6 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
7 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
8 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farlo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
9 (bases 1 to 115971)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J
```







MARKER CONFIRMATION: STSs D11S1310 and SHGC-19182 sequence confirmed  
MAPPED CLONE OVERLAP: pDJ1082L12 and 6-130a9.

FEATURES  
source

Location/Qualifiers  
1. .177142

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
complement(4265. .4373)  
/rpt\_family="MIR"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="MIR"  
75341. .75433  
/rpt\_family="MER44C"  
75448. .75858  
/rpt\_family="MER44C"  
76545. .76675  
/rpt\_family="MIR"  
complement(84972. .86432)  
/rpt\_family="MER7"  
complement(86435. .86629)  
/rpt\_family="Alu"  
complement(86632. .86977)  
/rpt\_family="THE1"  
complement(86978. .88519)  
/rpt\_family="MSTAR"  
complement(88552. .88906)  
/rpt\_family="THE1"  
complement(88908. .88988)  
/rpt\_family="Alu"  
89703. .89772  
/rpt\_family="MLT1"  
89833. .90211  
/rpt\_family="THE1"  
90212. .90583  
/rpt\_family="MLT1"  
91164. .91399  
/rpt\_family="Alu"  
91416. .91647  
/rpt\_family="L1"  
93141. .93267  
/rpt\_family="MIR"  
complement(94854. .95076)  
/rpt\_family="MIR"  
complement(97298. .97498)  
/rpt\_family="MIR"  
complement(100857. .101045)  
/rpt\_family="MIR"  
102052. .102354  
/rpt\_family="Alu"  
complement(103360. .103505)  
/rpt\_family="MIR"  
complement(105279. .105514)  
/rpt\_family="Alu"  
106255. .106562  
/rpt\_family="Alu"  
107761. .108120  
/rpt\_family="MER1"  
108430. .109760  
/rpt\_family="MER42"  
110182. .110221  
/rpt\_family="MER42"  
complement(110515. .110629)  
/rpt\_family="Alu"  
110649. .110926  
/rpt\_family="MER42"  
112736. .112780  
/rpt\_family="Alu"  
113214. .113313  
/rpt\_family="MIR"  
complement(114256. .114358)  
/rpt\_family="MIR"  
116565. .117029  
/rpt\_family="MLT1"  
117948. .118007  
/rpt\_family="MIR"  
118345. .118751  
/rpt\_family="THE1"  
complement(118866. .118996)

Query Match 87.2%; Score 21.8; DB 8; Length 177142;  
Best Local Similarity 92.0%; Pred. No. 18;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1  AAAAACTGCATCCCAATGTTATG 25
Db      2589 AAAAACTGCATCCCAATGTTATG 2613

RESULT 17
LOCUS   AC124078
DEFINITION Homo sapiens chromosome 11, clone RP1-59M18, complete sequence.
ACCESSION AC124078
VERSION   AC124078.5 GI:26080557
KEYWORDS HTG.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 177543)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
          Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Fitzgerald,M., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
          Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
          Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lindblad-Toh,K.,
          Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
          Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
          Meneus,L., Mihova,T., Mieng,A., Murphy,T., Naylor,J., Nguyen,C.,
          Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
          Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
          Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
          Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,
          Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
          Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 177543)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
          Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
          Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
          Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
          McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,A.,
          Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
          O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
          Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
          Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
          Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
          Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
          Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 177543)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
          Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
          Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
          Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
          McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,A.,
          Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
          O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
          Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
          Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
          Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
          Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
          Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (06-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Dec 6, 2002 this sequence version replaced gi:22549795.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27396
Center clone name: 59_M18
-----
FEATURES             Location/Qualifiers
     source            1..177543
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="11"
                        /map="11"
                        /clone="RP1-59M18"
                        /clone_lib="RPIC Human PAC library 1"
     repeat_region     1863..1886
                        /rpt_family="AT rich"
     repeat_region     complement(3634..3771)
                        /rpt_family="MIR3"
                        4136..4177
                        /note="single clone coverage"
                        complement(4256..4380)
                        /rpt_family="MIR"
                        4499..4581
                        /note="single clone coverage"
                        /rpt_family="L2"
                        5386..5528
                        /rpt_family="L2"
                        6596..6692
                        /rpt_family="MIR3"
                        8588..9349
                        /rpt_family="L2"
                        complement(9361..9434)
                        /rpt_family="MIR"
                        9746..9941
                        /rpt_family="AluSg/x"
                        complement(10213..10423)
                        /rpt_family="L1ME3"
                        complement(10510..10713)
                        /rpt_family="L1ME3"
                        complement(10757..11016)
                        /rpt_family="L1ME3"
                        complement(11029..11524)
                        /rpt_family="L1ME"
                        11530..11782
                        /rpt_family="MIR"
                        complement(12563..12712)
                        /rpt_family="MIR"
                        complement(12733..12808)
                        /rpt_family="MIR"

```



```

/clone="LJT24C17"
/clone_lib="LJT library"
/note="TAC clone:TW0406, synonym:Lotus japonicus"
4680. .4779
gap /estimated_length=unknown
6067. .6166
gap /estimated_length=unknown
8357. .8456
gap /estimated_length=unknown
11230. .11329
gap /estimated_length=unknown
14862. .14961
gap /estimated_length=unknown
17758. .17857
gap /estimated_length=unknown
22894. .22793
gap /estimated_length=unknown
27234. .27333
gap /estimated_length=unknown
31852. .31951
gap /estimated_length=unknown
38854. .38953
gap /estimated_length=unknown
45566. .45665
gap /estimated_length=unknown
52254. .52353
gap /estimated_length=unknown
58365. .58464
gap /estimated_length=unknown
66745. .66844
gap /estimated_length=unknown
84479. .84578
gap /estimated_length=unknown
108309. .108408
gap /estimated_length=unknown
ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 109919;
Best Local Similarity 91.7%; Pred. No. 55;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTCATCCCAATGTTATG 25
|||||
Db 60479 AAAAAGTCATCCCAATGTTATG 60502

RESULT 19
AR484315
LOCUS AR484315 1917 bp DNA linear PAT 14-MAY-2004
DEFINITION Sequence 1467 from patent US 6703492.
ACCESSION AR484315
VERSION AR484315.1 GI:47247418
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: US 6703492-A 1467 09-MAR-2004;
Smithline Beecham Corporation; Philadelphia, PA
FEATURES
source
1. .1917
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 81.6%; Score 20.4; DB 6; Length 1917;
Best Local Similarity 95.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAGTCATCCCAATGTTAT 24
|||||

```

```

Db 1441 AAAAAGTCATCCCAATGTTAT 1462

RESULT 20
AX142745
LOCUS AX142745 1917 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 1467 from Patent WO0134809.
ACCESSION AX142745
VERSION AX142745.1 GI:14282244
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 1467 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1. .1917
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="synthetic nucleic acid sequence"
ORIGIN
Query Match 81.6%; Score 20.4; DB 6; Length 1917;
Best Local Similarity 95.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAGTCATCCCAATGTTAT 24
|||||
Db 1441 AAAAAGTCATCCCAATGTTAT 1462

RESULT 21
AF270200
LOCUS AF270200 3276 bp DNA linear BCT 01-AUG-2000
DEFINITION Staphylococcus epidermidis strain SRI clone step.1053f03 genomic
sequence.
ACCESSION AF270200
VERSION AF270200.1 GI:9624109
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
AUTHORS Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc. 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source
1. .3276
/organism="Staphylococcus epidermidis"
/mol_type="genomic DNA"
/db_xref="taxon:1282"
/clone="step.1053f03"
ORIGIN

```

Query Match 81.6%; Score 20.4; DB 1; Length 3276;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24  
|||||  
Db 1193 AAAACTGCATCCTCAATGTTAT 1214

RESULT 22  
AR486154  
LOCUS AR486154 3276 bp DNA linear PAT 14-MAY-2004  
DEFINITION Sequence 4240 from patent US 6703492.  
ACCESSION AR486154  
VERSION AR486154.1 GI:47249257  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3276)  
Kimberly, W.J.  
AUTHORS Staphylococcus epidermidis nucleic acids and proteins  
TITLE Patent: US 6703492-A 4240 09-MAR-2004;  
JOURNAL SmithKline Beecham Corporation; Philadelphia, PA  
FEATURES  
source  
1..3276  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 81.6%; Score 20.4; DB 6; Length 3276;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24  
|||||  
Db 1193 AAAACTGCATCCTCAATGTTAT 1214

RESULT 23  
AX145518  
LOCUS AX145518 3276 bp DNA linear PAT 31-MAY-2001  
DEFINITION Sequence 4240 from Patent WO0134809.  
ACCESSION AX145518  
VERSION AX145518.1 GI:14284083  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
Kimberly, W.J.  
AUTHORS Staphylococcus epidermidis nucleic acids and proteins  
TITLE Patent: WO 0134809-A 4240 17-MAY-2001;  
JOURNAL GLAXO GROUP LIMITED (GB)  
FEATURES  
source  
1..3276  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/notes="synthetic nucleic acid sequence"

ORIGIN  
Query Match 81.6%; Score 20.4; DB 6; Length 3276;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24  
|||||  
Db 1193 AAAACTGCATCCTCAATGTTAT 1214

RESULT 24

CP000029.20/c

WFCOMMENT  
Sequence split into 27 fragments LOCUS CP000029 Accession CP000029

Fragment Name Begin End  
CP000029.00 1 110000  
CP000029.01 100001 210000  
CP000029.02 200001 310000  
CP000029.03 300001 410000  
CP000029.04 400001 510000  
CP000029.05 500001 610000  
CP000029.06 600001 710000  
CP000029.07 700001 810000  
CP000029.08 800001 910000  
CP000029.09 900001 1010000  
CP000029.10 1000001 1110000  
CP000029.11 1100001 1210000  
CP000029.12 1200001 1310000  
CP000029.13 1300001 1410000  
CP000029.14 1400001 1510000  
CP000029.15 1500001 1610000  
CP000029.16 1600001 1710000  
CP000029.17 1700001 1810000  
CP000029.18 1800001 1910000  
CP000029.19 1900001 2010000  
CP000029.20 2000001 2110000  
CP000029.21 2100001 2210000  
CP000029.22 2200001 2310000  
CP000029.23 2300001 2410000  
CP000029.24 2400001 2510000  
CP000029.25 2500001 2610000  
CP000029.26 2600001 2616530

Continuation (21 of 27) of CP000029 from base 2000001 (CP000029 Staphylococcus epidermidis

Query Match 81.6%; Score 20.4; DB 1; Length 110000;  
Best Local Similarity 95.5%; Pred. No. 85;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24  
|||||  
Db 92423 AAAACTGCATCCTCAATGTTAT 92402

RESULT 25  
AE016750/c  
LOCUS AE016750 300698 bp DNA linear BCT 01-JAN-2003

DEFINITION Staphylococcus epidermidis ATCC 12228, section 7 of 9 of the complete genome.  
ACCESSION AE016750 AE015929  
VERSION AE016750.1 GI:27316220

KEYWORDS Staphylococcus epidermidis ATCC 12228  
SOURCE Staphylococcus epidermidis ATCC 12228  
ORGANISM Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE 1 (bases 1 to 300698)  
Zhang, Y., Ren, S., Li, H., Fu, G., Lu, L., Lu, G., Jia, J., Tu, Y.,  
Qin, Z., Chen, Z. and Wen, Y.

TITLE Direct Submission  
JOURNAL Submitted (05-NOV-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China

FEATURES  
source  
1..300698  
/organism="Staphylococcus epidermidis ATCC 12228"  
/mol\_type="genomic DNA"  
/strain="ATCC 12228"  
/db\_xref="taxon:176280"  
complement(114..186)  
/product="tRNA-Lys"  
complement(190..264)  
/product="tRNA-Gln"  
complement(269..352)  
/product="tRNA-Tyr"  
complement(369..444)

```
/product="trNA-Val"
complement (451..522)
/product="trNA-Glu"
complement (524..598)
/product="trNA-Asn"
complement (612..726)
/product="5S ribosomal RNA"
complement (886..3713)
/product="23S ribosomal RNA"
complement (2718..2831)
/locus_tag="SE1755"
complement (2718..2831)
/locus_tag="SE1755"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO05396.1"
/db_xref="GI:27316221"
/translations="MMIIGRALFGRGAPLGNLPSDKLRMPINLTWSEHG"
complement (3977..5530)
/product="16S ribosomal RNA"
complement (5985..7052)
/locus_tag="SE1756"
complement (5985..7052)
/locus_tag="SE1756"
/codon_start=1
/transl_table=11
/product="Arp-binding Mrp-like protein"
/protein_id="AAO05397.1"
/db_xref="GI:27316222"
/translations="MLTNQVKEIVGALKDPIIDVLRSEGIIVSVIKDINHVSVK
VMAOLGGQQLDQWAIKALKENGANTVIRFELPSEVVERVYIGKSEKPTTEE
LLSQNPVEFISASGGVGKSTVAVNLAVALARSGKVGVLVDADIVGFSVDMGII
DERPGIDKEIIVRHGVKVIISMAFFVENAPVIRGPMGLKMLTNFTTEVQWGBLD
YLLDLDPGGDVLADVHSLPSSKEIIVTTPHTAAFAVARAGAMAKTEHTILGVI
ENMSYFESKTEKGVYFGKGGKLSDELTQLFAELPLEQPTWNPDPFSPSIYQSD
DRGELYTLARKVIASQKQ"
complement (7203..8642)
/locus_tag="SE1757"
complement (7203..8642)
/locus_tag="SE1757"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein"
/protein_id="AAO05398.1"
/db_xref="GI:27316223"
/translations="MASYELSKTKRNLIVSMILISAFVAILNQTLLNTALPHIMRELN
INENTSOMLTGFMLVNGVMIPLTAYLMDKVKTRPLYLGAMGSLIGSIVAALPNFG
VLMARVIOAGVAGILMPLMQFTLFTLFSKEHGFAMGLAGLVQPAPAGFTTGLI
IDNASWRPVPIVIGIALVFEIRGVSISSYNTKETLDRSVIYSTLGFGLMLYAF
SSAGNLGFSNPVLCVSLISLIIIGIFVKRQIISNPLLNKIFPNNKIFCESITSMI
IMLSWGPALLIPLYQNALGSLALLSGVLMPGAILINGMSVFTGPKFYDKYGRPLI
FTGPIILISCTFLLCFKVDTSYTLIVIAIRMAFVSLMMPINTTGINALKTEDIS
HGTAIMNFGVMAGSLCTALMTFMSICAQWVSSSESHASKEMIQRQSVAVGVDSFPA
LVTVFVILAFIALFIKEPNHLSNRKV"
complement (8912..9382)
/locus_tag="SE1758"
complement (8912..9382)
/locus_tag="SE1758"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO05399.1"
/db_xref="GI:27316224"
/translations="MMKFLNKSXHLVLTIVLITFVISGAILFTIFGFLYGLSLRIL
IYHLGDFSNKGYDNLIVYGVSIVLGYFTLFSIEHLMDFKKNLKNPFOGINFH
LISIVTIMPYFIVHLHYVHVNIHFVIMIIIGFLFVCKEVFPESKNLNKK"
complement (9570..10907)
/locus_tag="SE1759"
complement (9570..10907)
/locus_tag="SE1759"
/codon_start=1
/transl_table=11
/product="multidrug transporter"
/protein_id="AAO05400.1"
/db_xref="GI:27316225"
/translations="MNLKSIITVMALIIIMFMAAIETSIISLALPTIKNSLNAGNLVS
LVFTVVFIALVIANPIVGLMSRPKIIIVAVVGVLLFALGSLMSGLSOTFTFLIISRT
VQGFAGVMSLSQIVPKLAPEIPLRYKIMGVGVMSGSIIGPLGGAILEFASWH
WLFYINIFIAVAILLVLMTHFFPDETQVQSRPDIKGLIIFYIFIALMGFLGNQHH
IIFNIFSIILAVALLWLLFKIENSIEQPLPKFNFSILVIFVITDILLIAILMGYNL
YIPVYLOKLSLPSQSGFVFPPLSVAMITLNFNLAKEAHFTRKTLXICSPVLLVS
SLIMFGLKLLPLLIAFAVAVFAGLSFGYITKDSIVVQEEETSPKNMKOMMSFVALTKL
GSSVSGTIMGYMYALNVGLFGSNLNLVGLVLIIVAVCLIVMMNTLYKSTIQS"
complement (11146..11829)
/locus_tag="SE1760"
complement (11146..11829)
/locus_tag="SE1760"
/codon_start=1
/transl_table=11
/product="hemolysin III"
/protein_id="AAO05401.1"
/db_xref="GI:27316226"
/translations="MSQSKRKNDSSVETFKDIIPLTFGEIIGNAASHGAAALLTLFI
LPYAAVHSPNNGTLESISVSIVSIFPMFISSTIYHSMQNTSHKYLIRIDHSMI
VVAISGTVTPVLLSVVGGWGLVTLILMGTLWGLYKSIATKVNRLSLIYVLVWG
WVGIIFFLPIIIMRTSWWFFPFLGGLSYTIGANFVAQKNKPYFHMWHIFIVLASFL
HMGIFPYFM"
complement (11848..13035)
/locus_tag="SE1761"
complement (11848..13035)
/locus_tag="SE1761"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine pyrophosphorylase"
/protein_id="AAO05402.1"
/db_xref="GI:27316227"
/translations="MLDKNLEKYNQEHLYEYKELMSNEKNALDEKVDOLNLAIEOD
LQDLVFNKRTIDDVSSSEVKEYVKRNLNEERHTEYQKVEAIPNGEFAVLLMAGS
QTRQLYKGGPKGSFEIGTSLFELQARQIRLKEBTGHTINWTIMTSDINHKTIEYF
KQKIFYNDANHIHFQKQDNIVALSEEGKVLNRDGHIMETPNGNGVFSKSLKAGVY
DKMQDDHKVYIIFLNNIDNVKVLDPPLFAGFTVQSKDITSKTIQPKDSVSGRLNVV
DKDVTLEYSELDTDIVNQFNANIGIHAFKLGFTSAVDRELPHYHLAIKQLKLDEN
FGVVERPTLKFELFYFDIIFRYGTSFVTLQVPREEFSPLKNKEGKDSVHTATEDLKR
DLI"
complement (13090..13599)
/locus_tag="SE1762"
complement (13090..13599)
/locus_tag="SE1762"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO05403.1"
/db_xref="GI:27316228"
/translations="MTGKTHASCGFLVGAITTOYPHTDIFTSISVILSVISSILPDI
CHTQSKIGRRPLTSFPIRLGHRFTLSLFIIGISFLLYFIQTPMYVMAIVIGM
FSHVILDIITPKGVKLLYPLPFNIVSPIHFKTGGLVDVSLATALSVAIYTLQPQYLN
TMHWYWLK"
13776..14036
/locus_tag="SE1763"
13776..14036
/locus_tag="SE1763"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO05404.1"
/db_xref="GI:27316229"
/translations="MAWTVKQNDNEVRQIRVADIKIPNNEIKNVTQDDIHAVPEEN
GKEISIGTSFGKTNRLVLDTDQHLIITYQNDQKVNELTK"
14189..14503
/locus_tag="SE1764"
14189..14503
/locus_tag="SE1764"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
```

gene	/protein_id="AAO05405.1"	
	/db_xref="GI:27316230"	
	/translation="MFVVTNRITVKKGAKQWAPNFTKGGPIESLKGFEGLEVWQIDK	
	DDYSDMTVNSWTEEDFKWNVSDVFKQAHKNTGKSEDSFVIKSVKNSVLSLN	
CDS	RR"	
	complement (14682..16055)	
	/locus_tag="SE1765"	
	complement (14682..16055)	
gene	/locus_tag="SE1765"	
	/codon_start=1	
	/transl_table=11	
	/product="conserved hypothetical protein"	
gene	/protein_id="AAO05406.1"	
	/db_xref="GI:27316231"	
	/translation="WNGFKKVDIVGRKIFITMSLGLILFLVPIPVVDGKQQTLP	
	TAFLAGLMDGGIMPLIVITIVSGILITLCSTIYKKNLNPQGLSSAPNVKIGW	
gene	LIVRLAVFPFWMLEPTLPGRSTVDNLASFGIDGTGVLITSRQYGGYVSRREATV	
	LLELGPITFRPMTLFTPGRTVDNLASFGIDGTGVLITSRQYGGYVSRREATV	
	ISTTSFVSITFAIVAIETIRMQDOFFPYFLTVVISCLIAIMPRIWPLKNIPDEYA	
	KEYSEARNEQLPEKTKALKYGFDLATEVGIKSPGKFLSGFTVVDMMFVILPVV	
gene	MSIGTIATLIANTYTFVEIIGKPFVLELQIPEAHEASQIILGFADMFELSLIE	
	GVQNDVTRFVIGALSISQLVYLSEVGGVILGSKIPVSIKLFMIFLIRTIPLIIL	
	LAHLFIG"	
	complement (16352..17317)	
gene	/locus_tag="SE1766"	
	81.6%; Score 20.4; DB 1; Length 300698;	
	Query Match 95.5%; Pred. No. 80;	
	Best Local Similarity 0; Mismatches 1; Indels 0; Gaps 0;	
QY	Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
	3 AAAACTGCATCCCCAATGTTAT 24	
	290043 AAAACTGCATCTCAATGTTAT 290022	
LOCUS	BC099258	
	Xenopus laevis 1382 bp mRNA linear VRT 18-JUL-2005	
	IMAGE:6932197, partial cds.	
	BC099258	
KEYWORDS	BC099258.1 GI:68533884	
	Xenopus laevis	
	Xenopus laevis	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;	
	Xenopodinae; Xenopus; Xenopus.	
	1 (bases 1 to 1382)	
	Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.	
AUTHORS	and Richardson, P.	
	Genetic and genomic tools for Xenopus research: The NIH Xenopus	
	initiative	
	Dev. Dyn. 225 (4), 384-391 (2002)	
JOURNAL	1254917	
	2 (bases 1 to 1382)	
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,	
	Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,	
REFERENCE	Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,	
	Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,	
	Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,	
	Stapleton, J., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,	
AUTHORS	Schaeetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,	
	Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,	
	Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,	
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,	
JOURNAL	Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,	
	Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,	
	Fahey, J., Helton, E., Kettaman, M., Madan, A., Rodrigues, S.,	
	Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,	
REFERENCE	Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,	
	Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,	
	Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,	
JOURNAL	PUBLISHED	
	REFERENCE	
	Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.	
	Mammalian Gene Collection Program Team	
AUTHORS	Generation and initial analysis of more than 15,000 full-length	
	human and mouse cDNA sequences	
	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
	12477932	
JOURNAL	3 (bases 1 to 1382)	
	Klein, S. and Gerhard, D.S.	
	Direct Submission	
	Submitted (01-JUL-2005) National Institutes of Health, Xenopus Gene	
REFERENCE	Collection (XGC), National Institute of Child Health and Human	
	Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD	
	20892-7510, USA	
	NIH-MGC Project	
AUTHORS	Contact: XGC help desk	
	Email: c9apbs-remail.nih.gov	
	Tissue Procurement: Drs. Donald Brown and Lihuan Cai	
	cDNA Library Preparation: CLONTECH Laboratories, Inc.	
JOURNAL	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)	
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome	
	Center, Stanford University School of Medicine, Stanford, CA 94305	
	Web site: http://www-shgc.stanford.edu	
REFERENCE	Contact: (Dickson, Mark) mcd@paxil.stanford.edu	
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,	
	R. M.	
	Clone distribution: MGC clone distribution information can be found	
FEATURES	through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov	
	Series: IRAL Plate: 69 Row: c Column: 2	
	This clone was selected for full length sequencing because it	
	passed the following selection criteria: Hexamer frequency ORF	
source	analysis, similarity but not identity to protein.	
	Location/Qualifiers	
	1..1382	
	/organism="Xenopus laevis"	
gene	/mol_type="mRNA"	
	/db_xref="taxon:8355"	
	/clone="IMAGE:6932197"	
	/tissue_type="Whole, tadpole, metamorphosis, stage 62"	
CDS	/clone_lib="NICHD_XGC_Tag2"	
	/lab_host="DH10B"	
	/note="Vector: pDNR-LIB"	
	<1..1382	
CDS	/gene="LOC445845"	
	/db_xref="GeneID:445845"	
	<1..721	
	/gene="LOC445845"	
ORIGIN	/product="LOC445845 protein"	
	/protein_id="AAH99258.1"	
	/db_xref="GI:68533885"	
	/translation="VRGSRSLGTFCLARKNPKGMRPFITFLCFQAFILTVFSEE	
QY	KKTENEEOVKIEVLHLPSECSQKRGVLDVNAHYDAYLAQNMTKLSCSRATKDGHPK	
	WFLGVGVIVKGLDIALMGCSGKKKAIIPPSLAYGEAGHKIPNATLMFEIELYG	
	INFLGRSAEFYQIDLNDKLSKDEISHYLTEEFKQKQKQDPDSVHGILTDIPHN	
	DRDEDFISAKENVYGHDEL"	
Db	Query Match 80.8%; Score 20.2; DB 5; Length 1382;	
	Best Local Similarity 88.0%; Pred. No. 1.4e+02;	
	Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	1 AAAAACTCCATCCCCAATGTTATG 25	
RESULT 27		
	1286 AAAAACTGCATACAGATGTTATG 1310	
	BC077190	
	Xenopus laevis 1731 bp mRNA linear VRT 07-SEP-2004	
LOCUS	Xenopus laevis hypothetical protein LOC445845, mRNA (cDNA clone	
	IMAGE:4057027), partial cds.	
	DEFINITION	
	IMAGE:4057027), partial cds.	



ACCESSION	BC077190	1	GI:50416502	
VERSION	BC077190.1			
KEYWORDS	Xenopus laevis (African clawed frog)			
SOURCE	Xenopus laevis			
ORGANISM	Xenopus laevis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Xenopus.			
AUTHORS	1 (bases 1 to 1731)			
TITLE	Klein, S.L., Strauberg, R.L., Wagner, L., Pontius, J., Clifton, S.W. Genetic and genomic tools for Xenopus research: The NIH Xenopus Initiative			
JOURNAL	Dev. Dyn. 225 (4), 384-391 (2002)			
PUBMED	12454917			
REFERENCE	2 (bases 1 to 1731)			
AUTHORS	Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Strauberg, R.L., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.D., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
TITLE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
JOURNAL	12477932			
PUBMED	3 (bases 1 to 1731)			
REFERENCE	Klein, S. and Gerhard, D.S.			
AUTHORS	Direct Submission			
TITLE	Submitted (19-JUL-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA			
JOURNAL	NIH-MGC Project			
REMARK	Contact: XGC help desk			
COMMENT	Email: cgapbe-r@mail.nih.gov Tissue Procurement: Dr. Igor David cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca			
FEATURES	Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu, Parvaneh Saedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalios, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.			
source	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov Series: IRAC Plate: 149 Row: 0 Column: 24 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.			
	1. .1731 /organism="Xenopus laevis" /mol_type="mRNA"			

db_xref="taxon:8355"	
/clone="IMAGE:4057027"	
/issue type="Lung, adult Xenopus"	
/clone_lib="NICH_XGC_Lul"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
<1. .1731	
/gene="LOC445845"	
/db_xref="GeneID:445845"	
<1. .746	
/gene="LOC445845"	
/codon_start=3	
/product="LOC445845 protein"	
/protein_id="AAH77190.1"	
/db_xref="GI:50416503"	
/db_xref="GeneID:445845"	
/translation="TATCTIVTVGRSRSLGTFCCLARKKPKGMRFFITFLCQAFI LUTVSEEQTESEBEQVKIEVHLHPSECSQKSGDLVNAHYADLAQNTKLKSGR ATKDGHKFWFLVGQVIGKGLDIALMCMCSGEKKRAIIPPSLAYGAGHDKIPPNATL MFEILYGINRGRPSAEAFYQIDLNNKQSKDEISHYLTEFEEKDQKQDPVSHGTI LTDI1PHKDRDDEGDFISAEYNYVGHDEL"	
ORIGIN	
Query Match	80.8%; Score 20.2; DB 5; Length 1731;
Best Local Similarity	88.0%; Pred. No. 1.4e+02;
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1 AAAAACTGCATCCCAATGTTATG 25
Db	1311 AAAAACTGCATACAGATGTTATG 1335
RESULT 28	
AL732290	
LOCUS	AL732290 Human DNA sequence from clone CTD-2338C19 on chromosome X, complete sequence.
DEFINITION	AL732290.4 GI:21531515
ACCESSION	AL732290
VERSION	HTG.
KEYWORDS	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 66122)
AUTHORS	Whitehead, S.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT	On Jun 21, 2002 this sequence version replaced gi:21217881. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX CTD-2338C19 is from the Caltech genomic sperm BAC library D. VECTOR: pBelobAC11
	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: vegas@sanger.ac.uk ----- This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

source  
Location/Qualifiers  
1..66122  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="CTD-2338C19"  
/clone\_lib="CIT-HSP-D1"

ORIGIN

Query Match 79.2%; Score 19.8; DB 8; Length 66122;  
Best Local Similarity 91.3%; Pred. No. 1.7e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 3 AAAACTGCATCCCAATGTTATG 25  
||||| ||||| ||||| |||||  
Db 14935 AAAACTGAATCCCAATATTATG 14957

RESULT 29

AL355882 148468 bp DNA linear HTG 31-JAN-2002  
LOCUS Homo sapiens chromosome X clone RP3-479K6, 14 unordered pieces.  
DEFINITION AL355882  
ACCESSION AL355882  
VERSION AL355882.4 GI:9909022  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

REFERENCE

1  
McLay, K.  
Direct Submission  
Submitted (21-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Aug 24, 2000 this sequence version replaced gi:9797390.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: dj1185F7  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: pBlasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 140943 bases at least Q40  
Consensus quality: 143751 bases at least Q30  
Consensus quality: 145662 bases at least Q20  
Insert size: 147168; sum-of-contigs  
Insert size: 132352; 17.0% error; agarose-fp  
Quality coverage: 3.82x in Q20 bases; sum-of-contigs Quality  
coverage: 4.24x in Q20 bases; agarose-fp  
-----

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 3189: contig of 3189 bp in length  
\* 3190 3289: gap of 100 bp  
\* 3290 9680: contig of 6391 bp in length  
\* 9681 9780: gap of 100 bp  
\* 9781 12692: contig of 2912 bp in length  
\* 12693 12792: gap of 100 bp

\* 12793 19581: contig of 6789 bp in length  
\* 19582 19681: gap of 100 bp  
\* 19682 36219: contig of 16538 bp in length  
\* 36220 36319: gap of 100 bp  
\* 36320 46785: contig of 10466 bp in length  
\* 46786 46885: gap of 100 bp  
\* 46886 62406: contig of 15521 bp in length  
\* 62407 62507: gap of 100 bp  
\* 62508 97792: contig of 35286 bp in length  
\* 97793 97892: gap of 100 bp  
\* 97893 109555: contig of 11663 bp in length  
\* 109556 109655: gap of 100 bp  
\* 109656 112376: contig of 2721 bp in length  
\* 112377 112477: gap of 100 bp  
\* 112478 119534: contig of 7058 bp in length  
\* 119535 119635: gap of 100 bp  
\* 119636 123375: contig of 3741 bp in length  
\* 123376 123476: gap of 100 bp  
\* 123477 141241: contig of 17766 bp in length  
\* 141242 141341: gap of 100 bp  
\* 141342 148468: contig of 7127 bp in length.

FEATURES

source  
Location/Qualifiers  
1..148468  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP3-479K6"  
/clone\_lib="RPCL-3"  
1..3189  
/note="assembly fragment:00217  
fragment chain:1"  
3290..9680  
/note="assembly fragment:00951  
fragment chain:1"  
clone\_end:T7  
vector\_side:left"  
9781..12692  
/note="assembly fragment:00104  
fragment chain:2"  
12793..19581  
/note="assembly fragment:00296  
fragment chain:2"  
19682..36219  
/note="assembly fragment:00323"  
36320..46785  
/note="assembly fragment:00381"  
46886..62406  
/note="assembly fragment:00449"  
62507..97792  
/note="assembly fragment:00590"  
97893..109555  
/note="assembly fragment:00631"  
109656..112376  
/note="assembly fragment:00720"  
112477..119534  
/note="assembly fragment:00986"  
119635..123375  
/note="assembly fragment:01008"  
123476..141241  
/note="assembly fragment:01141"  
141342..148468  
/note="assembly fragment:01102  
clone\_end:SP6  
vector\_side:right"

ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 148468;  
Best Local Similarity 91.3%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 3 AAAACTGCATCCCAATGTTATG 25  
||||| ||||| ||||| |||||

Db 81449 AAAAAGTGAATCCCAATATTATG 81471

RESULT 30  
AC079592/c  
LOCUS  
DEFINITION  
AC079592 148497 bp DNA linear HTG 08-JAN-2003  
Homo sapiens chromosome 3 clone RP11-75415, WORKING DRAFT SEQUENCE,  
25 unordered pieces.  
AC079592.8 GI:20335660  
HTG; HTGS PHASE1; HTGS DRAFT.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 148497)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homesi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Lozato,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 148497)  
Worley,K.C.  
Direct Submission  
Submitted (04-SEP-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 148497)  
Worley,K.C.  
Direct Submission  
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Apr 28, 2002 this sequence version replaced gi:18449584.  
----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HBJF  
Center clone name: RP11-75415  
----- Summary Statistics  
Sequencing vector: M13;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 121512 bases at least Q40  
Consensus quality: 129899 bases at least Q30  
Consensus quality: 135781 bases at least Q20  
Estimated insert size: 144188; sum-of-contigs estimation  
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2325: contig of 2325 bp in length  
\* 2326 2425: gap of unknown length  
\* 2426 4893: contig of 2468 bp in length  
\* 4894 4894: gap of unknown length  
\* 4894 8632: contig of 3639 bp in length  
\* 8633 8732: gap of unknown length  
\* 8733 11069: contig of 2337 bp in length  
\* 11070 11189: gap of unknown length  
\* 11189 13945: contig of 2776 bp in length  
\* 13945 14045: gap of unknown length  
\* 14046 17046: contig of 3001 bp in length  
\* 17047 17146: gap of unknown length  
\* 17147 19481: contig of 2335 bp in length  
\* 19482 19581: gap of unknown length  
\* 19582 22448: contig of 2867 bp in length  
\* 22449 22548: gap of unknown length  
\* 22549 25891: contig of 3343 bp in length  
\* 25892 25991: gap of unknown length  
\* 25992 30431: contig of 4440 bp in length  
\* 30432 30531: gap of unknown length  
\* 30532 34280: contig of 3749 bp in length  
\* 34281 34380: gap of unknown length  
\* 34381 37568: contig of 3188 bp in length  
\* 37569 43299: contig of 5631 bp in length  
\* 43300 43399: gap of unknown length  
\* 43400 50957: contig of 7558 bp in length  
\* 50958 51057: gap of unknown length  
\* 51058 57906: contig of 6849 bp in length  
\* 57907 58006: gap of unknown length  
\* 58007 63428: contig of 5422 bp in length  
\* 63429 63528: gap of unknown length  
\* 63529 69051: contig of 5523 bp in length  
\* 69052 76910: contig of 7759 bp in length  
\* 76911 77010: gap of unknown length  
\* 77011 83900: contig of 6890 bp in length  
\* 83901 84000: gap of unknown length  
\* 84001 91544: contig of 7544 bp in length  
\* 91545 91644: gap of unknown length  
\* 91645 101701: contig of 10057 bp in length  
\* 101702 101801: gap of unknown length  
\* 101802 110735: contig of 8934 bp in length  
\* 110736 110835: gap of unknown length  
\* 110836 123862: contig of 13027 bp in length  
\* 123863 123962: gap of unknown length



Center clone name: CH230-292B24  
----- Summary Statistics  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 142822 bases at least Q40  
Consensus quality: 147110 bases at least Q30  
Consensus quality: 150075 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.bjsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.bjsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 75 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1330: contig of 1330 bp in length  
\* 1331 1430: gap of unknown length  
\* 1431 2800: contig of 1270 bp in length  
\* 2701 2800: gap of unknown length  
\* 2801 3940: contig of 1140 bp in length  
\* 3941 4040: gap of unknown length  
\* 4041 5156: contig of 1116 bp in length  
\* 5157 5256: gap of unknown length  
\* 5257 6859: contig of 1603 bp in length  
\* 6860 6959: gap of unknown length  
\* 6960 8417: contig of 1457 bp in length  
\* 8417 8516: gap of unknown length  
\* 8517 9808: contig of 1292 bp in length  
\* 9809 9909: gap of unknown length  
\* 9909 11040: contig of 1131 bp in length  
\* 11040 11139: gap of unknown length  
\* 11139 12169: contig of 1030 bp in length  
\* 12170 12269: gap of unknown length  
\* 12270 14012: contig of 1743 bp in length  
\* 14013 14112: gap of unknown length  
\* 14113 15810: contig of 1698 bp in length  
\* 15811 15910: gap of unknown length  
\* 15911 17216: contig of 1306 bp in length  
\* 17217 17316: gap of unknown length  
\* 17317 19178: contig of 1862 bp in length  
\* 19179 19278: gap of unknown length  
\* 19279 20521: contig of 1243 bp in length  
\* 20522 21622: gap of unknown length  
\* 21623 21722: contig of 1001 bp in length  
\* 21723 22909: contig of 1187 bp in length  
\* 22910 23010: gap of unknown length  
\* 23010 25670: contig of 2661 bp in length  
\* 25671 25770: gap of unknown length  
\* 25771 27137: contig of 1367 bp in length  
\* 27138 27237: gap of unknown length  
\* 27238 29209: contig of 1972 bp in length  
\* 29210 29309: gap of unknown length  
\* 29310 31130: contig of 1821 bp in length  
\* 31131 31230: gap of unknown length  
\* 31231 32812: contig of 1582 bp in length  
\* 32813 32912: gap of unknown length  
\* 32913 34538: contig of 1626 bp in length  
\* 34539 34638: gap of unknown length  
\* 34639 35647: contig of 1009 bp in length  
\* 35648 35747: gap of unknown length  
\* 35748 37822: contig of 2075 bp in length  
\* 37823 37922: gap of unknown length  
\* 37923 39275: contig of 1353 bp in length  
\* 39276 39375: gap of unknown length  
\* 39376 41609: contig of 2234 bp in length  
\* 41610 41709: gap of unknown length  
\* 41710 42936: contig of 1227 bp in length  
\* 42937 43036: gap of unknown length

\* 43037 44087: contig of 1051 bp in length  
\* 44088 44187: gap of unknown length  
\* 44188 45431: contig of 1244 bp in length  
\* 45432 45531: gap of unknown length  
\* 45532 47879: contig of 2348 bp in length  
\* 47880 47979: gap of unknown length  
\* 47980 49432: contig of 1453 bp in length  
\* 49433 49532: gap of unknown length  
\* 49533 52083: contig of 2551 bp in length  
\* 52084 52183: gap of unknown length  
\* 52184 53926: contig of 1743 bp in length  
\* 53927 54026: gap of unknown length  
\* 54027 56161: contig of 2135 bp in length  
\* 56162 56261: gap of unknown length  
\* 56262 57990: contig of 1729 bp in length  
\* 57991 58090: gap of unknown length  
\* 58091 60844: contig of 2754 bp in length  
\* 60845 60944: gap of unknown length  
\* 60945 63300: contig of 2356 bp in length  
\* 63301 63400: gap of unknown length  
\* 63401 64854: contig of 1454 bp in length  
\* 64855 64954: gap of unknown length  
\* 64955 67198: contig of 2244 bp in length  
\* 67199 67298: gap of unknown length  
\* 67299 69764: contig of 2466 bp in length  
\* 69765 69864: gap of unknown length  
\* 69865 72035: contig of 2171 bp in length  
\* 72036 72135: gap of unknown length  
\* 72136 74158: contig of 2023 bp in length  
\* 74159 74258: gap of unknown length  
\* 74259 76503: contig of 2245 bp in length  
\* 76504 76603: gap of unknown length  
\* 76604 79168: contig of 2565 bp in length  
\* 79169 79268: gap of unknown length  
\* 79269 81982: contig of 2714 bp in length  
\* 81983 82082: gap of unknown length  
\* 82083 84203: contig of 2121 bp in length  
\* 84204 84303: gap of unknown length  
\* 84304 87139: contig of 2836 bp in length  
\* 87140 87239: gap of unknown length  
\* 87240 89111: contig of 2472 bp in length  
\* 89112 89811: gap of unknown length  
\* 89812 92229: contig of 2418 bp in length  
\* 92230 92329: gap of unknown length  
\* 92330 94448: contig of 2119 bp in length  
\* 94449 94548: gap of unknown length  
\* 94549 96815: contig of 2267 bp in length  
\* 96816 96915: gap of unknown length  
\* 96916 100472: contig of 3556 bp in length  
\* 100472 100571: gap of unknown length  
\* 100572 103523: contig of 2952 bp in length  
\* 103524 103623: gap of unknown length

Query Match 79.2%; Score 19.8; DB 14; Length 185810;  
Best Local Similarity 91.3%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCATGTTA 23  
Db 94723 AAAAACTGCATCCCATGTTA 94745

RESULT 32  
CNS05TCF/c  
LOCUS 189389 bp DNA linear PRI 04-MAY-2001  
DEFINITION Human chromosome 14 DNA sequence BAC R-168L7 of library RCI-11  
from chromosome 14 of Homo sapiens (Human), complete sequence.  
ACCESSION AL355096 GI:13992190  
VERSION AL355096.4  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo
1 (bases 1 to 189389)
Heilig, R., Petit, J.-L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brotier, P., Catcolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 189389)
Genoscope.
Direct Submission
Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On May 8, 2001 this sequence version replaced gi:11876112.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-1029J19 (AC=AL049836)
Downstream BAC (overlapping the SP6 end) : C-2561P5 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 8.44x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : Bases
0 : 3
1 - 9 : 14
10 - 19 : 60
20 - 29 : 176
30 - 39 : 699
40 - 49 : 4072
50 - 59 : 8246
60 - 69 : 7529
70 - 79 : 13558
80 - 89 : 37982
90 - 99 : 117050
-----
Percentage of bases with a quality value >= 40 : 99 %.
-----
FEATURES
source
Location/Qualifiers
1. 189389
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RP11-11"
/clone="R-168L7"
/STS
67559..67694
/notes="matching EMBL:AI005270
Rdb:RH102424
dbSTS:STS69961
Identified using the e-PCR software (G. Schuler)"
ORIGIN
Query Match 79.2%; Score 19.8; DB 8; Length 189389;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTA 23
|||||
Db 82138 AACAACCTGCACCCCAATGTTA 82116
|||||

RESULT 33
AC112763 211635 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-246P24, *** SEQUENCING IN PROGRESS
DEFINITION

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

***, 7 unordered pieces.
AC112763
GI:25072605
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 211635)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Guarinate, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Khan, Z., King, L., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensu, L., Louis, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nair, L.,
Nwankwelu, O., Okwuonu, G., Olarnpung, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlarczyk, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 211635)
Worley, K. C.
Direct Submission
Submitted (24-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 211635)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```



## COMMENT

On Nov 19, 2002 this sequence version replaced gi:23195195.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GRXJ

Center clone name: CH230-246P24

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 197380 bases at least Q40

Consensus quality: 199965 bases at least Q30

Consensus quality: 201900 bases at least Q20

Estimated insert size: 199042; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 197215: contig of 197215 bp in length  
 \* 197216 197315: gap of unknown length  
 \* 197316 201157: contig of 3842 bp in length  
 \* 201158 201257: gap of unknown length  
 \* 201258 204646: contig of 3389 bp in length  
 \* 204647 204746: gap of unknown length  
 \* 204747 205898: contig of 1152 bp in length  
 \* 205899 207611: contig of 1613 bp in length  
 \* 207612 209032: contig of 1321 bp in length  
 \* 209033 209132: gap of unknown length  
 \* 209133 211635: contig of 2503 bp in length.

## FEATURES

source : 1. 211635  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-246P24"

1. .2887

/note="wgs end extension  
 clone\_end:Sp6"  
 2938. .4634

misc\_feature

/note="wgs end extension  
 clone\_end:Sp6"  
 5539. .7053

misc\_feature

/note="wgs end extension  
 clone\_end:Sp6"  
 complement(8698. .9583)

misc\_feature

/note="clone boundary  
 clone\_end:Sp6  
 site:  
 end sequence:BZ168459"  
 189679. .190526

misc\_feature

/note="clone boundary  
 clone\_end:T7  
 site:  
 end sequence:BZ168458"  
 191302. .193241  
 /note="wgs end extension  
 clone\_end:T7"  
 194084. .195139  
 /note="wgs end extension  
 clone\_end:T7"  
 195560. .197215  
 /note="wgs end extension  
 clone\_end:T7"  
 197216. .197315  
 /estimated length=unknown  
 197316. .198590  
 /note="wgs end extension  
 clone\_end:T7"  
 201158. .201257  
 /estimated length=unknown  
 201258. .202440  
 /note="wgs end extension  
 clone\_end:T7"  
 204647. .204746  
 /estimated length=unknown  
 205899. .205998  
 /estimated length=unknown  
 207612. .207711  
 /estimated length=unknown  
 209033. .209132  
 /estimated length=unknown

ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 211635;  
 Best Local Similarity 91.3%; Pred. No. 1.6e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAATGCATCCCCCAATGTTA 23  
 |||||  
 Db 116307 AAAAAAATGCATCCCCCACTTTA 116329

RESULT 34  
 BC045302/c 2124 bp mRNA linear VRT 08-MAR-2005  
 LOCUS  
 DEFINITION  
 MGC:55304 IMAGE:2600344, complete cds.  
 BC045302  
 ACCESSION  
 VERSION  
 BC045302.1 GI:28279506  
 KEYWORDS  
 MGC.  
 SOURCE  
 Danio rerio (zebrafish)  
 ORGANISM  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 2124)  
 Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,  
 Collins FS, Wagner L, Shenmen CM, Schuler GD, Altshul SF, Zeeberg  
 B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,  
 Max SI, Wang J, Heide F, Diatchenko L, Marusina K, Farmer AA, Rubin  
 GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,  
 Scheetz TE, Brownstein MJ, Ustin TB, Toshiyuki S, Carninci P,  
 Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullihy  
 SJ, Bosak SA, McSwan PJ, McKernan KJ, Malek JA, Gunaratne PH,  
 Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,  
 Villalón DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,  
 Helton E, Kettman M, Madan A, Rodrigues S, Sanchez A, Whiting M,  
 Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,  
 Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,  
 Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,  
 Smalusz DE, Schnerch A, Schein JE, Jones SJ and Marra MA.  
 Mammalian Gene Collection Program Team  
 Generation and initial analysis of more than 15,000 full-length

CONSRTM  
 TITLE







misc\_feature  
75582..85230  
/notes="assembly\_fragment:00174  
fragment\_chain:1"  
85331..91591  
/notes="assembly\_fragment:00023  
fragment\_chain:1"  
91692..113878  
/notes="assembly\_fragment:00773  
fragment\_chain:1  
clone\_end:77  
vector\_side:right"  
113979..135260  
/notes="assembly\_fragment:00469  
fragment\_chain:2"  
135361..140962  
/notes="assembly\_fragment:00089  
fragment\_chain:2"

ORIGIN

Query Match 77.6%; Score 19.4; DB 14; Length 140962;  
Best Local Similarity 95.2%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGT 21  
|||||  
Db 86821 AAAAACTGCATCCCAATGT 86841

RESULT 38

AL060513

LOCUS

DEFINITION

AL060513 172243 bp DNA linear ROD 17-DEC-2004  
Mouse DNA sequence from clone RP23-277B9 on chromosome 13 Contains  
the the Viri5, Viri14, Viri10, Viri2, Virh5 and Virh6 genes for  
vomeronasal 1 receptor 15, I14, I10, I2, H5 and H6, two Virh and  
two Viri pseudogenes and an ATP-binding cassette sub-family F  
(GCM20) member 2 (Abcf2) pseudogene, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL060513.3 GI:18476767  
HTG; Abcf2; Virh5; Viri10; Viri14; Viri2; Viri5.  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 172243)  
Lovell, J.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (16-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk  
Clone requests: Clonerequest@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:18181711.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
En:, EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep -----  
Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: http://mrcseq.har.mrc.ac.uk  
Contact: mouseq@har.mrc.ac.uk

-----  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.  
RP23-277B9 is from the RPCI-23 Mouse BAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6.

FEATURES  
source

1..172243  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/clone="RP23-277B9"  
/clone\_lib="RPCI-23"  
complement(20246..21142)  
/gene="Viri5"  
/locus\_tag="RP23-277B9.7-001"  
complement(20246..21142)  
/gene="Viri5"  
/locus\_tag="RP23-277B9.7-001"  
complement(20246..21142)  
/gene="Viri5"  
/locus\_tag="RP23-277B9.7-001"  
complement(20246..21142)  
/standard\_name="OTTMUSP00000000510"  
/note="match: proteins: Q8R261"  
/codon\_start=1  
/protein\_id="CAI24853.1"  
/db\_xref="GI:56205067"  
/translation="MNKITQAIVFLSLAGPGVGVNLFVRYVYTSALGTEKRPIDLI  
LIHLASNLIIICSTGVDTIVTVFVFRNFDGICGMVYLARWARGLSICTTCLLSV  
QVATISPTTIWTKLPQTSQVLPFLLELFWINVLISNLLSYIKAGSSINRSVAA  
TFIGHCYMLPSRHIIKWLFLSLMTLRDVIPOSMLGMSGSMALHLYKHKHVLVLRSS  
RFANSPPEIRATWSVLILMAFLFFYVWDEILSFYTGFTVTHDSTLLNIQKVLFEFGY  
ASFSPVLSRDVHPNVLHAH"  
complement(31384..33122)  
/locus\_tag="RP23-277B9.2-001"  
/pseudo  
complement(31384..33122)  
/locus\_tag="RP23-277B9.2-001"  
/note="match: proteins: Q8R271 Q8R272 Q8R280 Q8R281  
Q8R274 Q8R275 Q8R276 Q8R268 Q8R279"  
/pseudo  
/codon\_start=1  
42897..43793  
/gene="Virh14"  
/locus\_tag="RP23-277B9.11-001"  
42897..43793  
/gene="Virh14"  
/locus\_tag="RP23-277B9.11-001"  
42897..43793  
/gene="Virh14"  
/locus\_tag="RP23-277B9.11-001"  
/note="match: proteins: Q8R268 Q8R278 Q8R270 Q8R280  
Q8R272"  
/codon\_start=1  
/protein\_id="CAI24855.1"  
/db\_xref="GI:56205068"  
/translation="MVLQFIKETIPLFMTVMGILGNMVSVHYMSWGWGSPKKPIHL  
ILTLISFTNIILLAKGKQKTIIVGLRNFLLDDIGCKTIIVYLERVARGLSICTTCLLS  
VVOAIIISPRASGWSRLPKGAWHILPFFSFWILNGLISMNLIHSITSTGLNMSQLN  
NSKNKYCFMPRESREIKWIVLPLMVLRDVAVFOGAMGAGSCHMIFLLHKHQHVLVLRSS  
KLLYTPPELRAQSLLMLLCFFVFWYTDCAISLFLSLGLDSSSLMINIQKVLTLGY  
AVFSPVLVLRDGLLPACWHAQ"  
complement(52854..53331)  
/locus\_tag="RP23-277B9.5-001"  
/pseudo  
complement(52854..53331)  
/locus\_tag="RP23-277B9.5-001"  
/note="match: proteins: Q8R265 Q8R266 Q8R274 Q8R275  
Q8R276 Q8R277 Q8R279 Q8R272"  
/pseudo  
/codon\_start=1  
complement(53705..54173)



```
* 94584 185950: contig of 91367 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..185950
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="RP11-321D23"
        /clone_lib="RPC1-11.2"
        /note="assembly_fragment:01327
clone_end:SP6
vector_side:left"
misc_feature
  69905..86631
    /note="assembly_fragment:03228
fragment_chain:1"
misc_feature
  86732..94483
    /note="assembly_fragment:03627
fragment_chain:1"
misc_feature
  94584..185950
    /note="assembly_fragment:01854
fragment_chain:1
clone_end:T7
vector_side:right"
ORIGIN
  Query Match      77.6%; Score 19.4; DB 14; Length 185950;
  Best Local Similarity 95.2%; Pred. No. 2.5e+02;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 AAAAACTGCATCCCAATGTT 22
    |||||
Db   95613 AAAAACTGCATCCCAATATT 95593

RESULT 40
CR936412/c
LOCUS
DEFINITION
  CR936412.10 GI:71142827
  HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
  Danio rerio (zebrafish)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
  1 (bases 1 to 228881)
  Leongamornliert,D.
  Direct Submission
  Submitted (28-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  zfsh-help@sanger.ac.uk
  http://www.sanger.ac.uk/Projects/D_rerio/fags.shtml#dataeight
  On Jul 25, 2005 this sequence version replaced gi:66392990.
COMMENT
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: zfsh-help@sanger.ac.uk
  ----- Project Information
  Center project name: zK181M21
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Chemistry: Dye-terminator; 100% of reads
  Consensus quality: 227384 bases at least Q40
  Consensus quality: 227452 bases at least Q30
  Consensus quality: 227477 bases at least Q20
  Insert size: 228881; sum-of-contigs
  Insert size: 214479; 21.4% error; agarose-fp
  Quality coverage: 9.13x in Q20 bases; sum-of-contigs Quality
  coverage: 9.79x in Q20 bases; agarose-fp

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 228881: contig of 228881 bp in length.
  Location/Qualifiers
    1..228881
      /organism="Danio rerio"
      /mol_type="genomic DNA"
      /db_xref="taxon:9955"
      /chromosome="18"
      /clone="DKEY-181M21"
      /clone_lib="DanioKey"
  misc_feature
    1..228881
      /note="assembly_fragment:01220"
ORIGIN
  Query Match      77.6%; Score 19.4; DB 14; Length 228881;
  Best Local Similarity 95.2%; Pred. No. 2.4e+02;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 AAAAACTGCATCCCAATGTT 22
    |||||
Db   37197 AGAAACTGCATCCCAATGTT 37177

RESULT 41
AY598742
LOCUS
DEFINITION
  AY598742
  Trilobodrilus axi cytochrome c oxidase subunit I gene, partial cds;
  mitochondrial.
ACCESSION
  AY598742
VERSION
  AY598742.1 GI:51317447
KEYWORDS
  SOURCE
  ORGANISM
    mitochondrion Trilobodrilus axi
    Trilobodrilus axi
    Eukaryota; Metazoa; Annelida; Polychaeta; Scoleleida; Dinophillida;
    Dinophillidae; Trilobodrilus.
    1 (bases 1 to 479)
  Struck,T.H. and Purschke,G.
  The sister group relationship of Aeolosomatidae and Potamodrilidae
  (Annelida, 'Polychaeta') - a molecular phylogenetic approach based
  on 18S rDNA and Cytochrome Oxidase I
  Zool. Anz. 243, 281-293 (2005)
  2 (bases 1 to 479)
  Struck,T.H. and Purschke,G.
  Direct Submission
  Submitted (13-APR-2004) Life Science Department, Auburn University,
  College of Mathematics and Science, 101 Rouse Building, Auburn, AL
  36849, USA
  Location/Qualifiers
    1..479
      /organism="Trilobodrilus axi"
      /organelle="mitochondrion"
      /mol_type="genomic DNA"
      /db_xref="taxon:188474"
      /country="Germany: List, North Sea island of Sylt"
      <1..>479
        /codon_start=1
        /transl_table=5
        /product="cytochrome c oxidase subunit I"
        /protein_id="AA099901.1"
        /db_xref="GI:51317448"
  CDS
    1..479
      /translation="VWAXXMFVGMVDVTRATMIIVPTGKIVSWLATINGS
      PKMTEAPMLWAVGFIFLFTMGTLTGILANSSIDVLHDITYVVAHFHVLGSMGAVFA
      MFAGETHWPFPLFTGVNMHPRWTKAHFYMMFLGNLTFFPQHFGLGSLGMPRRYSYDPA
      "
```



to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
1..959  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..>959

## STS

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 959;  
Best Local Similarity 87.5%; Pred. No. 4.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 397 AAAAACTGCATATCCATTGTTATG 374

## RESULT 44

## BV543020

## LOCUS

DEFINITION BV543020 988 bp DNA linear STS 09-APR-2005  
spi43b08.g1 Clint Pan troglodytes verus STS genomic, sequence  
tagged site.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Pan troglodytes verus  
Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Pan.

## REFERENCE

## AUTHORS

## TITLE

Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome  
Unpublished (2005)

## JOURNAL

## COMMENT

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 988  
Protocol:  
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI  
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred

score >= 20, at least 30% of its base calls must satisfy SNOS(30,25)(single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

## source

1..988  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..>988

## STS

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 988;  
Best Local Similarity 87.5%; Pred. No. 4.4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 416 AAAAACTGCATATCCATTGTTATG 439

## RESULT 45

## AB122062/c

## LOCUS

## DEFINITION

AB122062 3079 bp mRNA linear INV 13-APR-2004  
Crassostrea gigas HSP68 mRNA for 68kDa heat shock protein, complete cds.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AB122062.1 GI:46359611  
Crassostrea gigas (Pacific oyster)  
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Ostreoida;  
Ostreidae; Ostreidae; Crassostrea.

## REFERENCE

## AUTHORS

## TITLE

Cloning and characterization of an inducible 68 kDa heat shock protein (HSP68) of the Japanese oyster (Crassostrea gigas)

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

Submitted (09-OCT-2003) Yoshihiro Yokoyama, Fukui Prefectural University, Department of Marine Bioscience, Gakuen-cyo 1-1, Obama, Fukui 917-0003, Japan (E-mail:yokoyama@pu.ac.jp, Tel:81770526300, Fax:81770526003)

## FEATURES

## source

1..3079  
/organism="Crassostrea gigas"  
/mol\_type="mRNA"  
/db\_xref="taxon:29159"  
/country="Japan"  
.3079  
/gene="HSP68"  
57..1898  
/gene="HSP68"  
/function="inducible chaperone"

## gene

## CDS



AUTHORS  
TITLE  
JOURNAL

Waterston, R.  
Direct Submission  
Submitted (19-NOV-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
8 (bases 1 to 39499)

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (16-MAY-2003) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
9 (bases 1 to 39499)

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (21-SEP-2004) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## COMMENT

Waterston, R.  
Direct Submission  
Submitted (21-SEP-2004) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

For a graphical representation of this clone sequence and its  
analysis see:  
http://www.wormbase.org/db/seq/sequence?name=F20B6;class=Sequence

## NEIGHBORING CLONE INFORMATION

The 5' clone is ZK470, 200 bp overlap; the 3' clone is W01H2, 1600  
bp overlap. Actual start of this clone is at base position 41972 of  
ZK470; actual end is at 38099 of F20B6.

## NOTES:

Coding sequences below are the result of integration and manual  
review of the following data : computer analysis using the program  
GeneFinder (P. Green and L. Hillier, personal communication), the  
large scale EST projects of Yuji Kohara  
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE INDEX.html) and The C.  
elegans ORFome cloning project (http://wofdb.dfci.harvard.edu/),  
similarity to other proteins from BlastX analyses  
(http://blast.wustl.edu/), sequence conservation with C. briggsae  
using Jim Kent's WABA alignment program (Genome Research  
10:1115-1125, 2000), individual C. elegans GenBank submissions,  
and personal communications with C. elegans researchers. tRNAs  
are predicted using the program tRNAscan-SE (Lowe, T.M. and  
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES  
source

1. .39499  
Location/Qualifiers  
/organism="Caenorhabditis elegans"  
/mol\_type="genomic DNA"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="X"  
/clone="F20B6"  
complement(385..8603)

gene

/note="HSP70 family"  
/codon\_start=1  
/product="68kDa heat shock protein"  
/protein\_id="BAD15285.1"  
/db\_xref="GI:46359612"

/translation="MEGRKPAVAGIDLGTSFCVGVFNRKGVDIINQDQNRTPPSVVA  
FVTEGLIGDGAKSQVTMAGNTVFDVKELIGRFTDESVMKHPKVVDDGKPK  
KVEVEYKQITRTPEIISAVLAKMKTAEPLGRVQEAIVTPAYFNNAQRATK  
DAGTLAGLVLRITNEPTAAALAFGLEKNIIGKKNVYDVGGLTDFDILITIDGVS  
FEVLSTAGTQIGGEDFDNLMTWTFVEEPRKKGKNIISGNSRLRLKTCACRKRIL  
SSSETSIDLDYIGEDIFYSLKRVITFELCSKJFFKTELEPKKALFAELPSQIQ  
EVLVGGSTRIPKIQITLQNMFKVNLNSVNIIDEAVYGAQAAILKGRSGIVQD  
MLRDNNVGTFFELNGIPPAFCGVQIENVDFDANGIVNVVTDRTDGTGSNLTWVKD  
KLSPDELNMKNKAMKYQBEDELEGORVEKHLSEYLIFFVOKCAQAFGEALDRELQ  
SLSLCNKTFSLWLDHNSGALFELEKLEVEQVFCSPIIKIQNASK"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 3079;  
Best Local Similarity 87.5%; Pred. No. 4.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACTGCATCCCAATGTTATG 25  
||||| ||||| ||||| ||||| |||||  
Db 1820 AAAAAAGTCAGCCCACTGTTATG 1797

## RESULT 46

U41015/c  
LOCUS 39499 bp DNA linear INV 21-SEP-2004  
DEFINITION Caenorhabditis elegans cosmid F20B6, complete sequence.

ACCESSION U41015

VERSION U41015.2 GI:21450561

KEYWORDS HTG.

SOURCE Caenorhabditis elegans

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 39499)

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## AUTHORS

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## TITLE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## JOURNAL

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## PUBMED

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## AUTHORS

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## TITLE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## JOURNAL

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## AUTHORS

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## TITLE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## JOURNAL

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## AUTHORS

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## TITLE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## JOURNAL

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

## REFERENCE

Waterston, R.  
Direct Submission  
Submitted (22-MAY-2002) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
On Jun 18, 2002 this sequence version replaced gi:1086643.  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

```
/gene="mrp-6"
/locus_tag="F20B6.3"
complement(join(385..561,611..886,1271..886,1271..1519,1565..1768,
2171..2274,2349..2469,2899..3005,3057..3207,3263..3342,
3195..3666,3713..4219,5271..6124,6179..6329,7017..7142,
7188..7563,7624..7826,8371..8603))
/gene="mrp-6"
/locus_tag="F20B6.3"
/standard_name="F20B6.3"
/note="contains similarity to Pfam domains PF00664 (ABC
transporter transmembrane region), PF00005 (ABC
transporter); coded for by the following C. elegans cDNAs:
yk297e4.5, yk326h11.5, yk605h4.5, yk679d5.5, yk297e4.3,
yk605h4.3, yk679d5.3"
/codon_start=1
/product="Multidrug resistance protein family protein 6"
/protein_id="AAA82317.2"
/db_xref="GI:30794620"
/db_xref="WormBase:F20B6.3"
/translation="MIPHQKTPDETKPDVNLESVEGVHPSEDSAGLISLPSFVG
FYWIRNAQTDLDLEKPSGISAKYAAQKLESWISSGPKRTHFLVGVHCTKVSU
MAHSLCLEELTRVAQPLLRLMSYDIEEGFFYPWPAMQVALIISLSLFSFTVCI
HHPYFHGLKVGNAHVATGLVLLYKSLRISLTSFETNSGQIIQLLNTDAKLRQAF
LFHVWMLCPLMFPFYAVILMSMGFCCLLGLFLVMAIVSFQVHYSYOLGVNRRNIGD
KNDRLKVMSEIINGIKVIMKMGWENAPAEKIRILREDEIKAVGTSAFVSSVMVGIFY
ISAKLGVYIFACDYDFEILLTRGVFPASSALFNSVRLLPFALFLPLGLLFGRELFTS
RRIDFFLCLEYSRTLELSDSTIDEDGKNEKEGTLKKIGEKPSWTLKCDRAEKITSV
FLNKVATVYDEHDQKKKFKALSDNDATVAGEKKDEPLHVYAVODLSPEFY
AGTCYIGISGVSGKSLILLTILSRMTRGNSLVSFPAYCAQEPWIPGSIQDNIV
FSGEKNERYENALCILLTPDLROPANGDMTIVGDSGTLSSGGQARVALARAVYAD
ADYLLDDPLSADVAHVGKLYENLRGLSKLVLVTHQIHFLSELSVILMKNNK
MYGNSDLKSTFTDDFKLOAVETEVSQDELFDVGGPPSPPPSVNLESMEKINKE
SIFQNSRLSKGSKSETYKREIVTESRKSGLTISWVYWEYIRGVINPLKGIWILV
CILTQLMNNFVDMWLNKWLVDLMNADNLSKEFNRRNDVLGTYFNVTQTYELT
FYIITVMSILGVWSVPRNAQLMASRILHNSMDFCVLKTESFEFDKNSEGMILNR
SDIGLTDLLAFTVFVFTGLTFVIGLVSMVRPIVILACAFITAGFYFCRMYV
HRSRELKTEAARSPLNTITSTVHGLSTRAYKENEMIEKFCALHDVNVAAVNMG
LLSARWFGICIDFLSVFSVIAIVVLQYETLTVGEVGLCLVCAVLQSGFSPSWLRQ
SAELNMGVSVSEVLEYTKGKEEARRPHDALDKQLNLSWPTKEIEFENVSVMYQ
ETVKNLGVSVLVGQKIGVGTGAGKSTLKIIFGLKEYCSGRVTIDGVLDHLSL
NFRRGMSIIPQEPVIFSGTVRENIDPDVEDDATWFALEQCELRKSITKEKGLNVE
LGQGVLSVSGROLPLFCARIMHRSKILVIDEATNVGDVTDACIQKTIRDFFKNAT
IITVAHLHTIMNSDNVMVFEKGLVEFDKPIITLLDDQSSLFSRMAAQSGAENFDSL
AALAHAIIDQSSYDTS"
13248..13249
/notes="Sli trans-splice acceptor; see yk759b01.5"
13252..14874
/gene="vha-12"
/locus_tag="F20B6.2"
/join(13252..13488,13538..13807,13859..14687,14735..14874)
/gene="vha-12"
/locus_tag="F20B6.2"
/standard_name="F20B6.2"
/note="contains similarity to Pfam domains PF02874 (ATP
synthase alpha/beta family, beta-barrel domain), PF00006
(ATP synthase alpha/beta family, nucleotide-binding
domain), PF00306 (ATP synthase alpha/beta chain, C
terminal domain); coded for by the following C. elegans
cDNAs: CB104433, OSTF003D11.1, OSTR003D11.1, yk86h1.5,
yk102c11.3, yk102c11.5, yk213a8.3, yk213a8.5, yk214g9.5,
yk239c12.3, yk243g10.5, yk291a2.5, yk304a10.3, yk304a10.5,
yk337d9.5, yk339b7.5, yk367d3.5, yk372a8.3, yk372a8.5,
yk386b10.3, yk386b10.5, yk448a3.5, yk448g12.5, yk453h4.5,
yk464h1.5, yk494h10.5, yk498h3.3, yk498h3.5, yk536a4.5,
yk628h3.3, yk628h3.5, yk633a3.5, yk640f5.5, yk647e12.5,
yk652a10.3, yk652a10.5, yk668h12.5, yk669e11.5,
yk759b01.5, yk759b01.5, yk777f09.3, yk777f09.5,
yk778d04.5, yk784e03.3, yk784e03.5, yk785h09.3,
yk785h09.5, yk786h08.3, yk786h08.5, yk790b12.3,
yk790b12.5, yk799c12.3, yk799c12.5, yk800d03.3,
yk800d03.5, yk836a11.3, yk837g06.3, yk837g06.5,
yk840f01.3, yk840f01.5, yk848f01.3, yk848f01.5,
yk850c10.3, yk852h06.3, yk"
```

AP008209_048	4800001	4910000	AP008209_121	12100001	12210000
AP008209_049	4900001	5010000	AP008209_122	12200001	12310000
AP008209_050	5000001	5110000	AP008209_123	12300001	12410000
AP008209_051	5100001	5210000	AP008209_124	12400001	12510000
AP008209_052	5200001	5310000	AP008209_125	12500001	12610000
AP008209_053	5300001	5410000	AP008209_126	12600001	12710000
AP008209_054	5400001	5510000	AP008209_127	12700001	12810000
AP008209_055	5500001	5610000	AP008209_128	12800001	12910000
AP008209_056	5600001	5710000	AP008209_129	12900001	13010000
AP008209_057	5700001	5810000	AP008209_130	13000001	13110000
AP008209_058	5800001	5910000	AP008209_131	13100001	13210000
AP008209_059	5900001	6010000	AP008209_132	13200001	13310000
AP008209_060	6000001	6110000	AP008209_133	13300001	13410000
AP008209_061	6100001	6210000	AP008209_134	13400001	13510000
AP008209_062	6200001	6310000	AP008209_135	13500001	13610000
AP008209_063	6300001	6410000	AP008209_136	13600001	13710000
AP008209_064	6400001	6510000	AP008209_137	13700001	13810000
AP008209_065	6500001	6610000	AP008209_138	13800001	13910000
AP008209_066	6600001	6710000	AP008209_139	13900001	14010000
AP008209_067	6700001	6810000	AP008209_140	14000001	14110000
AP008209_068	6800001	6910000	AP008209_141	14100001	14210000
AP008209_069	6900001	7010000	AP008209_142	14200001	14310000
AP008209_070	7000001	7110000	AP008209_143	14300001	14410000
AP008209_071	7100001	7210000	AP008209_144	14400001	14510000
AP008209_072	7200001	7310000	AP008209_145	14500001	14610000
AP008209_073	7300001	7410000	AP008209_146	14600001	14710000
AP008209_074	7400001	7510000	AP008209_147	14700001	14810000
AP008209_075	7500001	7610000	AP008209_148	14800001	14910000
AP008209_076	7600001	7710000	AP008209_149	14900001	15010000
AP008209_077	7700001	7810000	AP008209_150	15000001	15110000
AP008209_078	7800001	7910000	AP008209_151	15100001	15210000
AP008209_079	7900001	8010000	AP008209_152	15200001	15310000
AP008209_080	8000001	8110000	AP008209_153	15300001	15410000
AP008209_081	8100001	8210000	AP008209_154	15400001	15510000
AP008209_082	8200001	8300000	AP008209_155	15500001	15610000
AP008209_083	8300001	8410000	AP008209_156	15600001	15710000
AP008209_084	8400001	8510000	AP008209_157	15700001	15810000
AP008209_085	8500001	8610000	AP008209_158	15800001	15910000
AP008209_086	8600001	8710000	AP008209_159	15900001	16010000
AP008209_087	8700001	8810000	AP008209_160	16000001	16110000
AP008209_088	8800001	8910000	AP008209_161	16100001	16210000
AP008209_089	8900001	9010000	AP008209_162	16200001	16310000
AP008209_090	9000001	9110000	AP008209_163	16300001	16410000
AP008209_091	9100001	9210000	AP008209_164	16400001	16510000
AP008209_092	9200001	9310000	AP008209_165	16500001	16610000
AP008209_093	9300001	9410000	AP008209_166	16600001	16710000
AP008209_094	9400001	9510000	AP008209_167	16700001	16810000
AP008209_095	9500001	9610000	AP008209_168	16800001	16910000
AP008209_096	9600001	9710000	AP008209_169	16900001	17010000
AP008209_097	9700001	9810000	AP008209_170	17000001	17110000
AP008209_098	9800001	9910000	AP008209_171	17100001	17210000
AP008209_099	9900001	10010000	AP008209_172	17200001	17310000
AP008209_100	10000001	10110000	AP008209_173	17300001	17410000
AP008209_101	10100001	10210000	AP008209_174	17400001	17510000
AP008209_102	10200001	10310000	AP008209_175	17500001	17610000
AP008209_103	10300001	10410000	AP008209_176	17600001	17710000
AP008209_104	10400001	10510000	AP008209_177	17700001	17810000
AP008209_105	10500001	10610000	AP008209_178	17800001	17910000
AP008209_106	10600001	10710000	AP008209_179	17900001	18010000
AP008209_107	10700001	10810000	AP008209_180	18000001	18110000
AP008209_108	10800001	10910000	AP008209_181	18100001	18210000
AP008209_109	10900001	11010000	AP008209_182	18200001	18310000
AP008209_110	11000001	11110000	AP008209_183	18300001	18410000
AP008209_111	11100001	11210000	AP008209_184	18400001	18510000
AP008209_112	11200001	11310000	AP008209_185	18500001	18610000
AP008209_113	11300001	11410000	AP008209_186	18600001	18710000
AP008209_114	11400001	11510000	AP008209_187	18700001	18810000
AP008209_115	11500001	11610000	AP008209_188	18800001	18910000
AP008209_116	11600001	11710000	AP008209_189	18900001	19010000
AP008209_117	11700001	11810000	AP008209_190	19000001	19110000
AP008209_118	11800001	11910000	AP008209_191	19100001	19210000
AP008209_119	11900001	12010000	AP008209_192	19200001	19310000
AP008209_120	12000001	12110000	AP008209_193	19300001	19410000

AP008209\_194 19400001 19510000  
AP008209\_195 19500001 19610000  
AP008209\_196 19600001 19710000

Query Match 76.8%; Score 19.2; DB 15; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
Db 37592 AACAACTGCATGCCCAATGTTT 37569

AC154085 155293 bp DNA linear HTG 11-FEB-2005  
Rhinolophus ferrumequinum clone VMRC7-100N14, WORKING DRAFT  
SEQUENCE, 3 ordered pieces.

AC154085  
AC154085.2 GI:59276134  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
Rhinolophus ferrumequinum (Greater horseshoe bat)  
Rhinolophus ferrumequinum  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Laurasiatheria; Chiroptera; Microchiroptera;  
Rhinolophidae; Rhinolophinae; Rhinolophus.

REFERENCE 1 (bases 1 to 155293)  
ANTONELLIS,A., AYELE,K., BENJAMIN,B., BLAKESLEY,R.W., BOAKYE,A.,  
BOUFFARD,G.G., BRINKLEY,C., BROOKS,S., CHU,G., COLEMAN,H.,  
ENGLE,J., GESTOLE,M., GUAN,X., GUPTA,J., GUTIERREZ,P., HAGHIGHI,P.,  
HAN,J., HANSEN,N., HO,S.-L., HU,P., HUNTER,G., HURLE,B., IDOL,J.R.,  
JONES,C., KWONG,P., LARIC,P., LARSON,S., LEE-LIN,S.-Q., LEGASPI,R.,  
MADDEN,M., MADURO,Q.L., MADURO,V.B., MARGULIES,E.H., MASIello,C.,  
MASKER,B., MCDOWELL,J., MULLIKIN,J.C., OESTREICHER,J.S., PARK,W.,  
PORTNOY,M.B., PRASAD,A., PURI,O., REDDIX-DUGUE,N., ROSAS,B.,  
SCHANDLER,K., SCHUELER,M.G., SISON,C., STANTRIPOPO,S., STEPHEN,E.,  
TAYE,A., THOMAS,J.W., THOMAS,P.J., TSAIPOURI,V., VOGT,J.L.,  
WETHERBY,K.D., YOUNG,A. and GREEN,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 155293)  
Green,E.D.  
Direct Submission  
Submitted (22-DEC-2004) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
3 (bases 1 to 155293)  
Green,E.D.  
Direct Submission  
Submitted (11-FEB-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
On Feb 11, 2005 this sequence version replaced gi:56787717.

----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: Gcy  
Center clone name: 100N14

The sequence data in this record represents an 'enhanced'  
version of a Phase 2 submission. Specifically, the indicated  
order and orientation of each sequence contig, has been  
established using one or more of the following: read-pair  
data from individual subclones, overlaps with neighboring  
clones, alignment with available reference sequence (e.g.,  
human), and/or confirmation by PCR testing. In addition,  
the sequence assembly is generally based on at least 8X average  
coverage in Q20 bases and has been reviewed to rule out  
gross misassemblies, the low-quality ends of sequence  
contigs have been trimmed away, and each base is associated  
with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 154395 bases at least Q40  
Consensus quality: 154656 bases at least Q30  
Consensus quality: 154772 bases at least Q20  
Insert size: 174000; agarose-fp  
Quality coverage: 13.02x in Q20 bases; agarose-fp  
Quality coverage: 14.60x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 98407: contig of 98407 bp in length  
\* 98408 98507: gap of unknown length  
\* 98508 152440: contig of 53933 bp in length  
\* 152441 152540: gap of unknown length  
\* 152541 155293: contig of 2753 bp in length.

FEATURES  
Location/Qualifiers  
1..155293  
/organism="Rhinolophus ferrumequinum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:59479"  
/clone="VMRC7-100N14"  
/clone\_lib="VMRC7"  
/note="BAC resource: http://bacpac.chori.org/"

misc\_feature  
1..98407  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
98408..98507  
/estimated\_length=unknown  
98508..152440  
/note="assembly\_fragment"  
152441..152540  
/estimated\_length=unknown  
152541..155293  
/note="assembly\_fragment  
clone\_end:other  
vector\_side:right"

gap  
98408..98507  
/estimated\_length=unknown

misc\_feature  
98508..152440  
/note="assembly\_fragment"  
152441..152540  
/estimated\_length=unknown

gap  
152441..152540  
/estimated\_length=unknown

misc\_feature  
152541..155293  
/note="assembly\_fragment  
clone\_end:other  
vector\_side:right"

ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 155293;  
Best Local Similarity 87.5%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTAT 25  
Db 103349 AAATACTGAATCCCAAGTTATG 103326

RESULT 49  
AL360267  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP11-342C20 on chromosome 13 Contains  
a RNA-binding protein LIN-28 pseudogene, a novel gene and the 3'  
end of a novel gene, complete sequence.

ACCESSION  
AL360267  
VERSION  
AL360267.10 GI:13751441  
KEYWORDS  
HTG; LIN-28; RNA-binding protein.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 157381)



Consensus quality: 157076 bases at least Q40  
Consensus quality: 160029 bases at least Q30  
Consensus quality: 161054 bases at least Q20  
Insert size: 158000; agarose-fp  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 193: contig of 193 bp in length  
\* 194 293: gap of 100 bp  
\* 294 1492: contig of 1199 bp in length  
\* 1493 1592: gap of 100 bp  
\* 1593 2731: contig of 1139 bp in length  
\* 2732 2831: gap of 100 bp  
\* 2832 6919: contig of 4088 bp in length  
\* 6920 7019: gap of 100 bp  
\* 7020 11262: contig of 4243 bp in length  
\* 11263 11362: gap of 100 bp  
\* 11363 23705: contig of 12343 bp in length  
\* 23706 23805: gap of 100 bp  
\* 23806 46137: contig of 22332 bp in length  
\* 46138 46237: gap of 100 bp  
\* 46238 66133: contig of 19896 bp in length  
\* 66134 66233: gap of 100 bp  
\* 66234 91719: contig of 25486 bp in length  
\* 91720 119835: contig of 28016 bp in length  
\* 119836 119935: gap of 100 bp  
\* 119936 163328: contig of 43393 bp in length.

FEATURES

source

1. .163328  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-23N4"  
/clone\_lib="RP11 Human Male BAC"

misc\_feature

1. .193  
/note="assembly\_fragment"  
clone end:T7  
vector side:right"

gap 194. .293

misc\_feature /estimated\_length=100

294. .1492 /note="assembly\_fragment"

gap 1493. .1592

misc\_feature /estimated\_length=100

1593. .2731 /note="assembly\_fragment"

gap 2732. .2831

misc\_feature /estimated\_length=100

2832. .6919 /note="assembly\_fragment"

gap 6920. .7019

misc\_feature /estimated\_length=100

7020. .11262 /note="assembly\_fragment"

gap 11263. .11362

misc\_feature /estimated\_length=100

11363. .23705 /note="assembly\_fragment"

gap 23706. .23805

misc\_feature /estimated\_length=100

23806. .46137 /note="assembly\_fragment"

gap 46138. .46237

misc\_feature /estimated\_length=100  
46238. .66133 /note="assembly\_fragment"  
gap 66134. .66233  
misc\_feature /estimated\_length=100  
66234. .91719 /note="assembly\_fragment"  
clone end:SP6  
vector side:right"  
91720. .91819  
misc\_feature /estimated\_length=100  
91820. .119835 /note="assembly\_fragment"  
gap 119836. .119935  
misc\_feature /estimated\_length=100  
119936. .163328 /note="assembly\_fragment"

ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 163328;  
Best Local Similarity 87.5%; Pred No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 135995 AAAAACTGCATATCCAGTGTATG 136018  
|||||

Search completed: February 3, 2006, 21:27:13

Job time : 883 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-25

Perfect score: 25

Sequence: 1 aaaaacgcaccccaatgttatg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21.\*  
1: Geneseqm1980s.\*  
2: Geneseqm1990s.\*  
3: Geneseqm2000s.\*  
4: Geneseqm2001as.\*  
5: Geneseqm2001bs.\*  
6: Geneseqm2002as.\*  
7: Geneseqm2002bs.\*  
8: Geneseqm2003as.\*  
9: Geneseqm2003bs.\*  
10: Geneseqm2003cs.\*  
11: Geneseqm2003ds.\*  
12: Geneseqm2004as.\*  
13: Geneseqm2004bs.\*  
14: Geneseqm2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	89.6	3977	10	AD86804
2	22.4	89.6	3977	11	ADV31040
3	21.8	87.2	270	10	ACA56412
4	21.8	87.2	270	12	AD156208
5	21.8	87.2	491	5	ABAI1028
6	21.8	87.2	798	5	ABAI1400
7	21.8	87.2	798	5	ABAI15361
8	21.8	87.2	798	5	ABAI15360
9	21.8	87.2	798	5	ABAI14401
10	20.4	81.6	1914	8	ACA46717
11	20.4	81.6	1917	4	AAH53037
12	20.4	81.6	1953	6	ABN92144
13	20.4	81.6	1953	13	ADS01836
14	20.4	81.6	3276	4	AAH54876
15	19.8	79.2	95050	14	ADX98574
16	19.2	76.8	179651	10	ADL13813
17	18.8	75.2	53552	5	AAS13655
18	18.6	74.4	168174	6	ABT11173
19	18.6	74.4	168273	6	ABT11114

20	18.4	73.6	110000	14	ABE42401_27
21	18.2	72.8	521	10	ADE57979
22	18.2	72.8	521	10	ADE57983
23	18.2	72.8	521	10	ADE60446
24	18.2	72.8	521	10	ADE60727
25	18.2	72.8	521	10	ADE60731
26	18.2	72.8	95883	11	ACN43954
27	18.2	72.8	152048	10	ADL13855
28	17.8	71.2	578	6	ABQ31618
29	17.8	71.2	578	6	ABQ31618
30	17.8	71.2	1083	8	ACA24290
31	17.8	71.2	110000	14	ABE3175_33
32	17.8	71.2	162147	14	ABE35716
33	17.6	70.4	461	9	ACH39621
34	17.6	70.4	549	4	AAH09944
35	17.6	70.4	596	4	AAS28698
36	17.6	70.4	596	10	ADG41894
37	17.6	70.4	596	11	ADI97668
38	17.6	70.4	612	5	ABV56236
39	17.6	70.4	614	4	AAS28126
40	17.6	70.4	614	10	ADG41030
41	17.6	70.4	614	11	ADI96804
42	17.6	70.4	1086	8	ACA53946
43	17.6	70.4	1506	8	ACA43061
44	17.6	70.4	1770	11	ACH94314
45	17.6	70.4	1803	11	ACH94414
46	17.6	70.4	1951	13	ADR60058
47	17.6	70.4	2975	10	ADB68941
48	17.6	70.4	3833	10	ADE25037
49	17.6	70.4	4363	4	AAK68667
50	17.6	70.4	4963	4	AAJ36538
51	17.6	70.4	4963	8	ABX59526
52	17.6	70.4	4963	12	ADJ30276
53	17.6	70.4	8423	5	ABA15397
54	17.6	70.4	10256	4	ABK42295
55	17.6	70.4	10256	9	ABQ60451
56	17.6	70.4	12563	6	ABQ79513
57	17.6	70.4	12563	12	ADQ19748
58	17.6	70.4	12563	13	ADQ89825
59	17.6	70.4	12563	14	AEA24024
60	17.6	70.4	12775	12	ADQ23893
61	17.6	70.4	13572	9	AAU58743
62	17.6	70.4	13819	5	ABA15398
63	17.6	70.4	15944	5	ABA15399
64	17.6	70.4	68200	14	ADX80722
65	17.6	70.4	110000	6	ABQ69245_06
66	17.6	70.4	110000	6	ABQ67197_05
67	17.6	70.4	110000	14	ABE39175_31
68	17.6	70.4	229301	14	ABE35719
69	17.4	69.6	2179	4	AAI58201
70	17.4	69.6	2179	5	ADQ98408
71	17.4	69.6	2179	9	ADBA48168
72	17.4	69.6	3708	6	ABA05449
73	17.4	69.6	3790	5	ADL46016
74	17.4	69.6	5546	10	ACC06673
75	17.4	69.6	5546	10	ADK64761
76	17.4	69.6	17528	6	ABL32601
77	17.2	68.8	427	13	ADP60633
78	17.2	68.8	681	8	ACA23085
79	17.2	68.8	866	2	AAT90428
80	17.2	68.8	1012	9	AAD56875
81	17.2	68.8	1875	12	ADJ39903
82	17.2	68.8	3198	9	ACD39241
83	17.2	68.8	3198	9	ACD39240
84	17.2	68.8	3198	9	ACD39243
85	17.2	68.8	3198	9	ACD39242
86	17.2	68.8	3917	2	AAT31723
87	17.2	68.8	7422	4	ABL08910
88	17.2	68.8	7447	4	ABL19158
89	17.2	68.8	8132	4	ABL19160
90	17.2	68.8	36714	13	ABD33628
91	17.2	68.8	36724	14	ADZ13602
92	17.2	68.8	38605	11	ACN44050

Continuation (28 o  
Ades57979 Human gen  
Ades57983 Human gen  
Ade60446 Human gen  
Ade60727 Human gen  
Ade60731 Human gen  
Adn43954 Human gen  
Adl13855 Osteoarth  
Abq31618 Oligonuc  
Aca24290 Prokaryot  
Continuation (34 o  
Aeb35716 L. pneumo  
Ach39621 Human foe  
Aah09944 Human cdn  
Aas28698 Genomic s  
Adg41894 Human res  
Adi97668 Human res  
Abv56236 Human pro  
Aas28126 Novel cdn  
Adg41030 Human res  
Adi96804 Human res  
Aca53946 Prokaryot  
Aca43061 Prokaryot  
Ach94314 Klebsiell  
Ach94414 Klebsiell  
Adh60058 Cotton cd  
Adh68941 C. neofo  
Ade25037 Plant gro  
Aak68667 Human imm  
Aal36538 Human mus  
Abx59526 cDNA enco  
Adj30276 Human mus  
Abal5397 Human ner  
Abk42295 Genomic s  
Adb60451 Connectiv  
Abq79513 TRRAP nuc  
Adq19748 Human sof  
Adq89825 Antagonis  
Aea24024 Human PRO  
Adg23893 Human sof  
Ad58743 Human tra  
Abal5398 Human ner  
Abal5399 Human ner  
Adx80722 Human man  
Continuation (7 o  
Continuation (6 of  
Continuation (32 o  
Aeb35719 L. pneumo  
Aai58201 Human pol  
Adq98408 DNA encod  
Adb48168 Novel hum  
Aba05449 Human tra  
Adl46016 Human ova  
Acc06673 Gene sequ  
Adk64761 Disease t  
Ab132601 Human imm  
Adr60633 Cotton cd  
Aca23085 Prokaryot  
Aat90428 K-ras Gen  
Aad56875 Human C-t  
Adj39903 Plant cdn  
Acd39241 Mosquito  
Acd39240 Mosquito  
Acd3243 Mosquito  
Acd3242 Mosquito  
Aat31723 K-ras onc  
Ab108910 Drosophil  
Ab119158 Drosophil  
Abd33628 Human can  
Adz13602 Human can  
Acn44050 Human gen



```
93 17.2 68.8 62231 11 ACN44176
94 17.2 68.8 86453 14 ADY25730
c 95 17.2 68.8 96588 9 ADA03026
c 96 17.2 68.8 96588 10 ADB72764
c 97 17.2 68.8 96588 10 ADC85506
c 98 17.2 68.8 96588 12 ADM74621
c 99 17.2 68.8 110000 2 AAX20248_04
c 100 17.2 68.8 110000 2 AAX20248_05
c 101 17.2 68.8 116277 2 AAX20249
c 102 17.2 68.8 177866 10 ADL13935
c 103 17.2 68.8 216215 10 ADF69167
c 104 17.2 68.8 264965 12 ADN16203
c 105 17.2 68.8 268685 6 ABS56563
c 106 17 68.0 134 3 AAC12915
c 107 17 68.0 171 3 AAC18716
c 108 17 68.0 243 10 ADH82265
c 109 17 68.0 317 13 ADU14854
c 110 17 68.0 426 14 ACL61932
c 111 17 68.0 495 10 ADK60905
c 112 17 68.0 529 5 ADL44781
c 113 17 68.0 613 5 ADL63641
c 114 17 68.0 654 10 ADF03192
c 115 17 68.0 660 14 AEA54313
c 116 17 68.0 697 13 ADQ54890
c 117 17 68.0 851 5 ADL38397
c 118 17 68.0 851 5 ADL73266
c 119 17 68.0 871 13 ADX28923
c 120 17 68.0 879 4 AA194139
c 121 17 68.0 918 9 ADA31381
c 122 17 68.0 1021 10 ADK58271
c 123 17 68.0 1173 13 ADX10353
c 124 17 68.0 1204 2 AAX13339
c 125 17 68.0 1204 6 ABS99134
c 126 17 68.0 1574 5 ABA16953
c 127 17 68.0 2000 13 ADX10967
c 128 17 68.0 2782 14 ADZ48893
c 129 17 68.0 3012 4 AAH98552
c 130 17 68.0 3089 4 ABL24022
c 131 17 68.0 3811 4 ABL24612
c 132 17 68.0 5582 14 ADX06207
c 133 17 68.0 5749 14 ADX06203
c 134 17 68.0 5759 14 ADX06205
c 135 17 68.0 5780 12 ADN03901
c 136 17 68.0 5780 13 ADP23421
c 137 17 68.0 5925 13 ADU05939
c 138 17 68.0 5927 11 ADN95930
c 139 17 68.0 5927 14 ADX06201
c 140 17 68.0 6916 4 AAK78177
c 141 17 68.0 9031 4 ABL24594
c 142 17 68.0 26636 4 ABL16650
c 143 17 68.0 35042 3 AAAB1454
c 144 17 68.0 45698 12 ADJ12490
c 145 17 68.0 107820 3 AAD16230
c 146 17 68.0 110000 3 AAAB1489_1
c 147 17 68.0 110000 6 ABQ74964_4
c 148 17 68.0 110000 6 ABQ74964_5
c 149 17 68.0 110000 14 AEA61169_0
c 150 17 68.0 130312 14 AEB39168
```

## ALIGNMENTS

```
RESULT 1
ADE86804
ID ADE86804 standard; cDNA; 3977 BP.
```

```
XX
AC ADE86804;
```

```
XX
DT 29-JAN-2004 (first entry)
```

```
XX
DE Kv3.1 alpha-subunit gene.
```

```
KW ss; gene; rat ; Kv3.1 ; alpha-subunit; voltage-gated potassium channel ;
KW GABAergic interneuron; brain; parvalbumin-containing interneuron;
KW repolarisation; fast-firing phenotype ; prefrontal cortex ;
KW schizophrenia; splice variant.
```

```
OS Rattus norvegicus.
```

```
XX Key Location/Qualifiers
FT CDS 1162..2919
FT /*tag= a
FT /product= "Kv3.1 alpha subunit"
```

```
PN EP1348963-A1.
```

```
XX 01-OCT-2003.
```

```
PF 28-MAR-2002; 2002EP-000071114.
```

```
PR 28-MAR-2002; 2002EP-000071114.
```

```
XX (MITS-) MITSUBISHI PHARMA CORP.
```

```
PI Cochran S, Yamagami K, Ohashi Y;
```

```
DR WPI: 2003-805901/76.
```

```
DR P-PSDB; ADE86805.
```

```
PT New polynucleotides and polypeptides, useful for diagnosing
PT schizophrenia, and for identifying a compound modulating the activity of
PT Kv3 channels.
```

```
XX Claim 1; Page 9-16; 24pp; English.
```

```
PS This sequence encodes the rat Kv3.1 alpha-subunit. Kv3.1 is a voltage-
gated potassium channel which is predominantly located within a subset of
gated potassium channels within the brain, namely the parvalbumin-
gated potassium channels. Its function is to allow fast repolarisation of
containing interneurons. The cells and thus, contributes to the fast-firing phenotype of these
neurons. Parvalbumin expression has been shown to be altered within the
prefrontal cortex in schizophrenia. There are two alternative splice
variants of the Kv3.1 alpha-subunit, Kv3.1a and Kv3.1b, with each variant
only differing at the intracellular C-terminal. The Kv3.1 alpha-subunit
polynucleotide and polypeptide are useful for diagnosing schizophrenia,
and for identifying compounds that modulate the expression of the
polypeptide or polynucleotide, or the activity of Kv3.1. They are also
useful for the identification of subjects who are predisposed to
schizophrenia, and for evaluating the efficacy of drugs for such
disorder, and monitoring the progress of patient symptoms involved in
clinical trials for the treatment of such disorder.
```

```
XX Sequence 3977 BP; 825 A; 1268 C; 1113 G; 771 T; 0 U; 0 Other;
```

```
XX Query Match 89.6%; Score 22.4; DB 10; Length 3977;
Best Local Similarity 95.8%; Pred. No. 3;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 AAAAACTGCATCCCAATGTTATG 25
```

```
Db 3915 AAAAACTGCATCCCAATGTTATG 3938
```

```
RESULT 2
```

```
ADY31040
```

```
XX ID ADY31040 standard; DNA; 3977 BP.
```

```
XX AC ADY31040;
```

```
XX DT 05-MAY-2005 (first entry)
```

```
XX DE Rat voltage-gated potassium channel Kv3.1 DNA.
```

```
XX diagnosis; schizophrenia; potassium channel; gene; ds.
```

```
XX
```

```

OS Rattus norvegicus.
XX
XX Key Location/Qualifiers
XX CDS 1162..2919
XX FT /*tag= a
XX FT /product= "voltage-gated potassium channel Kv3.1"
XX FT /gene= "Kcnc1"
XX
XX EP1348964-A2.
XX
XX 01-OCT-2003.
XX
XX 27-MAR-2003; 2003EP-00007002.
XX
XX 28-MAR-2002; 2002EP-00007114.
XX
XX (MITS-) MITSUBISHI PHARMA CORP.
XX
XX Cochran S, Yamagami K, Ohashi Y;
XX
XX WPI; 2003-805902/76.
XX P-PSDB; ADY31041.
XX REFSEQ; NM_012856.
XX
XX New polynucleotides and polypeptides, useful for diagnosing
XX schizophrenia, and for identifying a compound modulating the activity of
XX Kv3 channels.
XX
XX Claim 1; SEQ ID NO 1; 50pp; English.
XX
XX The invention relates to the use of a polynucleotide comprising a
XX sequence of 3977, 2441 or 3410 base pairs, fully defined in the
XX specification, its fragment or complement, for diagnosis or treating
XX schizophrenia. The invention also claims a method for screening for a
XX compound regulating the expression of the polynucleotide comprising: (a)
XX bringing a test compound into contact with a cell capable of expressing
XX the polynucleotide, whose expression is desired to be controlled; (b)
XX detecting the expression of the polypeptide in the cell; and (c)
XX determining a compound promoting or suppressing the expression of the
XX polypeptide compared to a control. The polynucleotides and polypeptides
XX are useful for diagnosing schizophrenia, and for identifying a compound
XX modulating the activity of Kv3 channels. They are also useful for the
XX identification of subjects who are predisposed to schizophrenia, and for
XX evaluating the efficacy of drugs for such disorder, and monitoring the
XX progress of patient symptoms involved in clinical trials for the
XX treatment of such disorder. This sequence corresponds to the gene for rat
XX voltage-gated potassium channel Kv3.1.
XX
XX Sequence 3977 BP; 825 A; 1268 C; 1113 G; 771 T; 0 U; 0 Other;
XX
XX Query Match 89.6%; Score 22.4; DB 11; Length 3977;
XX Best Local Similarity 95.8%; Pred. No. 3;
XX Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 2 AAAAAGTCATCCCAATGTTATG 25
XX |||||
XX Db 3915 AAAAAGTCATCCCAATGTTATG 3938
XX
XX RESULT 3
XX ID ACA56412 standard; cDNA; 270 BP.
XX AC ACA56412;
XX
XX 06-JUN-2003 (first entry)
XX
XX Norway rat signalling pathway polynucleotide probe SEQ ID NO 1010.
XX
XX Norway rat; probe; ss; array element; Parkinson's disease;
XX signalling pathway population; cancer; adenocarcinoma; leukaemia;
XX immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX

```

```

OS Rattus norvegicus.
XX
XX US6500938-B1.
XX
XX 31-DEC-2002.
XX
XX 30-JAN-1998; 98US-00016434.
XX
XX 30-JAN-1998; 98US-00016434.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Au-Young J, Seilhamer JJ;
XX
XX WPI; 2003-352189/33.
XX
XX Combination of polynucleotide probes, useful as array elements in a
XX microarray for monitoring the expression of a number of target
XX polynucleotides.
XX
XX Claim 1; SEQ ID NO 1010; 65pp; English.
XX
XX The invention relates to a combination which, comprises a number of
XX polynucleotide probes comprising a sequence selected from one of the 1490
XX sequences mentioned in the specification. The combination is useful as an
XX array element in a microarray for monitoring the expression of a number
XX of target polynucleotides. The microarray is particularly useful in the
XX diagnosis and treatment of cancer and immunopathology and neuropathology.
XX The microarray is useful in diagnostics and treatment regimens, drug
XX discovery and development, toxicological and carcinogenicity studies,
XX forensics and pharmacogenomics. The microarray is also useful for
XX monitoring progression of diseases and for developing sophisticated
XX profiles for the effects of currently available therapeutic drugs. The
XX combination is also useful for purifying a subpopulation of mRNAs, cDNAs
XX and genomic fragments and in research and diagnostic applications. The
XX array can detect changes in expression in a large number of genes coding
XX for different signalling pathway populations which can be used to diagnose
XX various diseases including cancer e.g. adenocarcinoma and leukaemia,
XX immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
XX and Parkinson's disease. The present sequence represents a polynucleotide
XX probe of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
XX Sequence 270 BP; 87 A; 56 C; 52 G; 75 T; 0 U; 0 Other;
XX
XX Query Match 87.2%; Score 21.8; DB 10; Length 270;
XX Best Local Similarity 92.0%; Pred. No. 3.9;
XX Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAGTCATCCCAATGTTATG 25
XX |||||
XX Db 164 AAAAAGTCATCCCAATGTTATG 188
XX
XX RESULT 4
XX ID ADI56208 standard; DNA; 270 BP.
XX AC ADI56208;
XX
XX 22-APR-2004 (first entry)
XX
XX Human polynucleotide probe #1010.
XX
XX Human; probe; ss; receptor-like polypeptide; transducing polypeptide;
XX effector-like polypeptide; cancer; immunopathology; neuropathology;
XX drug development; toxicology; carcinogenicity;
XX signalling pathway polypeptide; adrenal gland; bladder; bone;
XX bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;
XX diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
XX dementia; amnesia; epilepsy; Alzheimer's disease; depression.
XX

```

```
XX Homo sapiens.
OS
XX US2004010136-A1.
PN
XX
XX 15-JAN-2004.
PD
XX
XX 26-NOV-2002; 2002US-00305720.
PF
XX
XX 30-JAN-1998; 98US-00016434.
PR
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Au-Young J, Seilhamer JJ;
PI
XX WPI; 2004-090520/09.
XX
XX New composition comprising polynucleotide probes, useful as array
PT elements in a microarray for monitoring the expression of target
PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
PT fragments.
XX
XX Claim 6; SEQ ID NO 1010; 73bp; English.
XX
XX The invention relates to a composition of polynucleotide probes
CC comprising first polynucleotide probes comprising at least a portion of a
CC gene encoding a receptor-like polypeptide, second polynucleotide probes
CC comprising at least a portion of a gene encoding a transducing
CC polypeptide and third polynucleotide probes comprising at least a portion
CC of a gene encoding an effector-like polypeptide. The probes of the
CC composition are useful as array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray is useful in the
CC diagnosis and treatment of cancer, an immunopathology or a
CC neuropathology. It can also be used for drug discovery and development,
CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.
CC Microarrays can also be used for monitoring the progression of diseases
CC that may be associated with the altered expression of signalling pathway
CC polypeptides. The composition can also be used to purify a subpopulation
CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile
CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of
CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,
CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or
CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,
CC epilepsy, Alzheimer's disease or depression. This sequence represents a
CC human polynucleotide probe of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 270 BP; 87 A; 56 C; 52 G; 75 T; 0 U; 0 Other;
XX
Query Match 87.2%; Score 21.8; DB 12; Length 270;
Best Local Similarity 92.0%; Pred. No. 3.9;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAACTGCATCCCAATGTTATG 25
Db 164 AAAAACTGCATCCCAATGTTATG 188
RESULT 5
ABAl1028
ID ABAl1028 standard; cDNA; 491 BP.
XX
XX ABAl1028;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 35.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antiskickling; antianaemic; antiarthritic; cancer;
XX
```

```
KW antirheumatic; hepatotropic; cerebrotropic; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
```

PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 27-SEP-2000; 2000US-0236327P.  
 PR 28-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 20-OCT-2000; 2000US-0242221P.  
 PR 01-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249267P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251160P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-541565/60.  
 DR P-PSDB; ABB14702.  
 XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system cancers  
 and metastases.  
 XX Claim 1; SEQ ID NO 35; 1701pp + Sequence Listing; English.  
 XX The invention relates to novel genes (AB11004-ABA21534) and proteins  
 (AB114678-AB18001) useful for preventing, treating or ameliorating  
 medical conditions e.g. by protein or gene therapy. The genes are  
 isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 disorders e.g. Addison's disease, allergies, diabetes mellitus, Crohn's disease,  
 anaemia, autoimmune thyroiditis, rheumatoid arthritis and ulcerative colitis; (c)  
 multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (d) wound healing  
 cardiovascular disorders such as myocardial ischaemia; (e) neurological diseases  
 such as e.g. cerebral anoxia and epilepsy; and (f)  
 infectious diseases such as viral, bacterial, fungal and parasitic  
 infections. Note: The sequence data for this patent did not form part of  
 the printed specification, but was obtained in electronic format directly  
 from WIPO at ftp.wipo.int/pub/published\_pt\_sequences  
 XX Sequence 491 BP; 137 A; 120 C; 100 G; 132 T; 0 U; 2 Other;  
 SQ  
 Query Match 87.2%; Score 21.8; DB 5; Length 491;  
 Best Local Similarity 92.0%; Pred. No. 4.2;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAACTGCATCCCAATGTTATG 25  
 Db 447 AAAAACTGCATCCCAATGTTATG 471  
 RESULT 6  
 ABA14400  
 ID ABA14400 standard; DNA; 798 BP.  
 AC ABA14400;  
 XX 23-JAN-2002 (first entry)  
 DT Human nervous system related polynucleotide SEQ ID NO 6731.  
 DE Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX WO200159063-A2.  
 XX 16-AUG-2001.  
 PD 17-JAN-2001; 2001WO-US001334.  
 PF 31-JAN-2000; 2000US-0179065P.  
 PR

PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating nervous system cancers  
XX  
XX and metastases.  
XX

PS Disclosure; SEQ ID NO 6731; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins

CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy. The genes are

CC isolated from a range of human tissues disclosed in the specification.

CC The nucleic acids, proteins, antibodies and (a) antagonists are useful in

CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,

CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune

CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic

CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)

CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing

CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)

CC infectious diseases such as viral, bacterial, fungal and parasitic

CC infections. Note: The sequence data for this patent did not form part of

CC the printed specification, but was obtained in electronic format directly

CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

XX

SQ Sequence 798 BP; 228 A; 184 C; 173 G; 213 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 5; Length 798;

Best Local Similarity 92.0%; Pred. No. 4.5;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTTATG 25

Db 764 AAAAACTGCATCCCAATGTTATG 788

RESULT 7

ABAI5361

ID ABAI5361 standard; DNA; 798 BP.

XX

AC ABAI5361;

XX

DT 23-JAN-2002 (first entry)

DE

DE Human nervous system related polynucleotide SEQ ID NO 7692.

XX

KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;

KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;

KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;

KW antineumatic; hepatotropic; cerebroprotective; antinflammatory;

KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;

KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX

OS Homo sapiens.

XX

XX WO200159063-A2.

XX

XX 16-AUG-2001.

XX

XX 17-JAN-2001; 2001WO-US001334.

XX

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 05-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 21-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234597P.

PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239335P.

PR 13-OCT-2000; 2000US-023937P.

PR 20-OCT-2000; 2000US-0240360P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.

PR 20-OCT-2000; 2000US-0242221P.

PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 01-DEC-2000; 2000US-0251160P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 useful for preventing, diagnosing and/or treating nervous system cancers  
 and metastases.

Disclosure; SEQ ID NO 7692; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
 (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
 medical conditions e.g. by protein or gene therapy. The genes are  
 isolated from a range of human tissues disclosed in the specification.  
 The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,  
 breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
 ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)

CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 798 BP; 227 A; 184 C; 174 G; 213 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 5; Length 798;

Best Local Similarity 92.0%; Pred. No. 4.5;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAGTCATGCCCAATGTTATG 25

||||| ||||| ||||| ||||| |||||

Db 764 AAAAAGTCATGCCCAATGTTATG 788

RESULT 8

ABAI5360

ID ABAI5360 standard; DNA; 798 BP.

XX ABAI5360;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 7691.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antiskicking; antianaemic; antirheumatic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

PN WO200159063-A2.

XX 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US001334.

PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.







PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 01-DEC-2000; 2000US-0251160P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-541565/60.  
 XX  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system cancers  
 PT and metastases.  
 XX  
 PS Disclosure; SEQ ID NO 6732; 1701pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
 CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and antagonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 798 BP; 227 A; 184 C; 174 G; 213 T; 0 U; 0 Other;  
 Query Match 87.2%; Score 21.8; DB 5; Length 798;  
 Best Local Similarity 92.0%; Pred. No. 4.5;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 AAAAAGCTGCATCCCAATGTTATG 25  
 DB 764 AAAAAGCTGCATCCCAATGTTATG 788  
 RESULT 10  
 ACA46717  
 ID ACA46717 standard; DNA; 1914 BP.  
 XX  
 AC ACA46717;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #28374.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Staphylococcus epidermidis.

XX WO200277183-A2.  
 XX 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362899P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU42847.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 34587; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing of the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1914 BP; 648 A; 321 C; 323 G; 622 T; 0 U; 0 Other;  
 Query Match 81.6%; Score 20.4; DB 8; Length 1914;  
 Best Local Similarity 95.5%; Pred. No. 23;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 3 AAAAAGCTGCATCCCAATGTTAT 24  
 DB 1441 AAAAAGCTGCATCCCAATGTTAT 1462  
 RESULT 11  
 AAH53037  
 ID AAH53037 standard; DNA; 1917 BP.

```

XX AAH53037;
AC
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1467.
XX
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW endocarditis; ds.
XX
XX Staphylococcus epidermidis.
OS
XX
XX WO200134809-A2.
PN
XX
XX 17-MAY-2001.
PD
XX
XX 09-NOV-2000; 2000WO-US030782.
PF
XX
XX 09-NOV-1999; 99US-0164258P.
PR
XX
XX (GLAX ) GLAXO GROUP LTD.
PA
XX
XX Kimmerly WJ;
PI
XX
XX WPI; 2001-316495/33.
DR
XX
XX P-PSDB; AAG82187.
DR
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
PT
XX
XX Claim 8; Page 414-415; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
XX Sequence 1917 BP; 650 A; 321 C; 323 G; 623 T; 0 U; 0 Other;
SQ
Query Match 81.6%; Score 20.4; DB 4; Length 1917;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
Db 1441 AAAACTGCATCCCAATGTTAT 1462

RESULT 12
ABN92144
ID ABN92144 standard; DNA; 1953 BP.
AC
XX
XX ABN92144;
AC
XX
XX 24-JUL-2002 (first entry)
DT
XX
XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1607.
DE

```

```

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.
XX
XX Staphylococcus epidermidis.
OS
XX
XX US6380370-B1.
PN
XX
XX 30-APR-2002.
PD
XX
XX 13-AUG-1998; 98US-00134001.
XX
XX 14-AUG-1997; 97US-0055779P.
XX
XX 08-NOV-1997; 97US-0064964P.
PR
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Doucette-Stamm LA, Bush D;
PI
XX
XX WPI; 2002-381255/41.
DR
XX
XX P-PSDB; ABP39599.
DR
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
PT
XX
XX Disclosure; SEQ ID NO 1607; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
XX Sequence 1953 BP; 663 A; 320 C; 338 G; 632 T; 0 U; 0 Other;
SQ
Query Match 81.6%; Score 20.4; DB 6; Length 1953;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
Db 1477 AAAACTGCATCCCAATGTTAT 1498

RESULT 13
ADS01836
ID ADS01836 standard; DNA; 1953 BP.
XX
XX ADS01836;
AC
XX
XX 04-NOV-2004 (first entry)
DT
XX
XX Staphylococcus epidermis polynucleotide seqid 1131.
DE
XX
XX antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system; gene; ds.
XX
XX Staphylococcus epidermidis.
OS
XX
XX US2004147734-A1.
XX
XX 29-JUL-2004.
PD
XX
XX 01-DEC-2003; 2003US-00724972.
XX
XX 08-NOV-1997; 97US-0064964P.
XX
XX 13-AUG-1998; 98US-00134001.
PR

```

PR 29-NOV-1999; 99US-00450969.  
XX (DOUC/) DOUCETTE-STAMM L.  
PA (BUSH/) BUSH D.  
XX  
XX  
XX Doucette-Stamm L, Bush D;  
XX  
XX WPI; 2004-580138/56.  
XX P-PSDB; ADS05608.  
XX  
XX New isolated polypeptide and encoding nucleic acid derived from  
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
PT treating an S. epidermidis bacterial infection.  
XX  
XX Claim 5; SEQ ID NO 1131; 741pp; English.  
XX  
XX The invention describes an isolated nucleic acid comprising a nucleotide  
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
CC given in the specification. Also described are: a recombinant expression  
CC vector; a cell comprising a recombinant expression vector of (i);  
CC comprising an S. epidermidis polypeptide; an isolated nucleic acid  
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection, comprising a nucleic acid cited above and a carrier; treating  
CC a subject for S. epidermidis infection; a recombinant or substantially  
CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
CC sample; a computer readable medium having recorded in it the nucleotide  
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
CC system for identifying fragments of the Staphylococcus genome of  
CC commercial importance; a computer based system for identifying fragments  
CC of the Staphylococcus plasmids of commercial importance; identifying  
CC commercially important nucleic acid fragments of the Staphylococcus  
CC genome and/or plasmids; and identifying an expression modulating fragment  
CC of the Staphylococcus genome and/or plasmids. The methods and  
CC compositions of the present invention are useful for the diagnosis,  
CC prevention and/or treatment of an Staphylococcal epidermidis bacterial  
CC infection. This sequence encodes a S. epidermis protein of the invention.  
XX  
XX SQ Sequence 1953 BP; 663 A; 320 C; 338 G; 632 T; 0 U; 0 Other;  
  
Query Match 81.6%; Score 20.4; DB 13; Length 1953;  
Best Local Similarity 95.5%; Pred. No. 23;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 AAAACTGCATCCCAATGTTAT 24  
Db 1477 AAAACTGCATCCTCAATGTTAT 1498  
|||||  
  
RESULT 14  
AAH54876  
ID AAH54876 standard; DNA; 3276 BP.  
XX  
XX AAH54876;  
XX  
XX 03-SEP-2001 (first entry)  
XX  
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:4240.  
XX  
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;  
XX endocarditis; ds.  
XX  
XX Staphylococcus epidermidis.  
XX  
XX WO200134809-A2.  
XX  
XX 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US030782.

XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX  
XX Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
XX  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
XX Claim 8; Page 1960-1961; 2188pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
XX SQ Sequence 3276 BP; 1139 A; 525 C; 545 G; 1067 T; 0 U; 0 Other;  
  
Query Match 81.6%; Score 20.4; DB 4; Length 3276;  
Best Local Similarity 95.5%; Pred. No. 25;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 AAAACTGCATCCCAATGTTAT 24  
Db 1193 AAAACTGCATCCTCAATGTTAT 1214  
|||||  
  
RESULT 15  
ADX98574  
ID ADX98574 standard; DNA; 95050 BP.  
XX  
XX ADX98574;  
XX  
XX 05-MAY-2005 (first entry)  
XX  
XX Human LOC145197 genomic DNA.  
XX  
XX SNP detection; breast tumor; endocrine disease;  
XX gynecology and obstetrics; neoplasm; cytostatic; metastasis;  
XX gene therapy; RNA interference; ds; SNP; single nucleotide polymorphism;  
XX LOC145197.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX variation 225  
XX /\*tag= a  
XX /standard\_name= "Single nucleotide polymorphism (SNP) "  
XX variation 505  
XX /\*tag= b  
XX /standard\_name= "Single nucleotide polymorphism (SNP) "  
XX variation 5170  
XX /\*tag= c  
XX /standard\_name= "Single nucleotide polymorphism (SNP) "

```
FT variation
FT /tag= d
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 6839
FT /tag= e
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 7016
FT /tag= f
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 7237
FT /tag= g
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 8190
FT /tag= h
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 8357
FT /tag= i
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 8577
FT /tag= j
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 15999
FT /tag= k
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 17722
FT /tag= l
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 18543
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 21639
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 24753
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 26289
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 27120
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 28737
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 29562
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 30078
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 30974
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 31545
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 31779
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 34062
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 34831
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 34906
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 34934
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 34938
```

```
FT variation
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 35131
FT /tag= ac
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 35511
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 35708
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 36071
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 36824
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 36959
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 37335
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 37343
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38507
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38554
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38624
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38690
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38744
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 38756
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 39525
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 40682
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 41181
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 41414
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 41657
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 42821
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 45475
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 45611
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 45763
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 46038
FT /tag= az
```

FT variation /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 46344 /tag= ba  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 46669 /tag= bb  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 46926 /tag= bc  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 47069 /tag= bd  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 47293 /tag= be  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 47602 /tag= bf  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 4821 /tag= bg  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 49875 /tag= bh  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 51178 /tag= bi  
FT /standard\_name= "Single nucleotide polymorphism (SNP)"  
FT 51180

Query Match 79.2%; Score 19.8; DB 14; Length 95050;  
Best Local Similarity 91.3%; Pred. No. 78;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTA 23  
||| ||||| ||||| |||||  
Db 29898 AACAACTGCAACCCCAATGTTA 29920

RESULT 16  
ADL13813  
ID ADL13813 standard; DNA; 179651 BP.  
XX  
AC ADL13813;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Osteoarthritis-associated polymorphic nucleotide #345.  
XX  
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;  
KW joint space narrowing; osteophyte development; joint pain;  
KW osteoarthritis; SNP; single nucleotide polymorphism.  
XX  
OS Homo sapiens.  
XX  
PN WO2003054166-A2.  
XX  
PD 03-JUL-2003.  
XX  
PF 19-DEC-2002; 2002WO-US041225.  
XX  
PR 20-DEC-2001; 2001US-0342603P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jones KA, Schafer A;  
XX  
DR WPI; 2003-559141/52.  
XX  
PT Determining susceptibility of an individual to joint space narrowing,  
PT osteophyte development and/or joint pain comprises identifying whether  
PT the individual has at least one polymorphism in a polynucleotide encoding  
PT ..a protein.

XX Disclosure; SEQ ID NO 345; 297pp; English.  
XX  
CC The invention relates to a method of determining susceptibility of an  
CC individual to joint space narrowing and/or osteophyte development and/or  
CC joint pain comprising identifying whether the individual has at least one  
CC polymorphism in a polynucleotide encoding at least one of the protein  
CC listed in the specification. The methods, composition and agent are  
CC useful for modulating the susceptibility of an individual to joint space  
CC narrowing and/or osteophyte development and/or joint pain that is  
CC associated with a disease, preferably osteoarthritis. The cell line and  
CC the non-human animal are useful for screening for an agent for diagnosing  
CC an individual having susceptibility to joint space narrowing and/or  
CC osteophyte development and/or joint pain. This sequence corresponds to  
CC the polynucleotide encoding a protein listed in the specification. (Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences).  
XX  
SQ Sequence 179651 BP; 48335 A; 39400 C; 42185 G; 49731 T; 0 U; 0 Other;  
Query Match 76.8%; Score 19.2; DB 10; Length 179651;  
Best Local Similarity 87.5%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACTGCATCCCAATGTTATG 25  
||| ||||| ||||| |||||  
Db 106059 AGAAATGAATCCCAATGTTATG 106082  
RESULT 17  
AAS13655  
ID AAS13655 standard; DNA; 53552 BP.  
XX  
AC AAS13655;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Genomic DNA sequence encoding human hspG15.  
XX  
KW Human; reproductive-specific protein; male infertility; gene therapy;  
KW spermatogenesis; sperm count disorder; anti infertility; reproduction;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200166752-A2.  
XX  
PD 13-SEP-2001.  
XX  
PF 07-MAR-2001; 2001WO-US007371.  
XX  
PR 07-MAR-2000; 2000US-0187518P.  
PR 12-JAN-2001; 2001US-0261557P.  
XX  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX  
PI Wang PJ, Page DC;  
XX  
DR WPI; 2001-570774/64.  
XX  
PT Novel reproduction-specific protein, useful for treating disorders of  
PT reduced sperm count, enhancing/increasing sperm count and/or sperm  
PT activity.  
XX  
PS Claim 1; Fig 61; 151pp; English.  
XX  
CC The present invention relates to the isolation of novel mammalian and  
CC human reproductive-specific proteins (AAU07859-AAU07899), and the nucleic  
CC acids encoding them. The nucleic acids encoding reproductive-specific  
CC proteins are useful for diagnosing infertility which is a result of  
CC reduced sperm count, reduced sperm motility, malformed sperm or  
CC combinations of these. The sequences of the invention are useful as



CC markers for spermatogonial cells, for identifying genes or proteins  
CC characteristic of male infertility, diagnosing or aiding in the diagnosis  
CC of infertility in men, and for contraception in which sperm production or  
CC sperm count is reduced or defective sperm is produced. Antibodies to  
CC reproductive-specific proteins are useful for determining the presence of  
CC these proteins in a sample obtained from a man being assessed for  
CC infertility, for identifying the expression of genes in particular cell  
CC type or particular developmental stage, for studies of spermatogenesis,  
CC and for immunofluorescence of germ cells or in Western blots for  
CC assessing the presence of the protein the antibody binds. The sequences  
CC of the invention are also useful for treating disorders of reduced sperm  
CC count, and for increasing sperm count and/or sperm activity. The nucleic  
CC acids of the invention are useful in gene therapy. AAS13648-AAS13671  
CC represent nucleic acid sequences encoding for the human reproduction-  
CC specific proteins of the present invention

XX  
SQ Sequence 53552 BP; 14743 A; 13420 C; 12475 G; 12914 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 5; Length 53552;  
Best Local Similarity 90.9%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ABACTGCATCCCAATGTTATG 25  
||||| ||||||| |||||  
Db 29520 AACTTCATCCCAATGTTATG 29541

RESULT 18  
ABT11173  
ID ABT11173 standard; DNA; 168174 BP.  
XX  
AC ABT11173;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 63.  
XX  
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;  
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;  
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;  
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;  
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;  
ds.

XX Homo sapiens.  
XX WO200262825-A2.  
XX 15-AUG-2002.  
XX 07-FEB-2002; 2002WO-US003546.  
XX  
XX 08-FEB-2001; 2001US-0267515P.  
XX 21-AUG-2001; 2001US-0314248P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Barnes G, Meyer J;  
XX WPI; 2002-627522/67.  
XX

XX New isolated nucleic acid molecule with an allelic variant of a  
XX polymorphic region of an 5-LO gene, useful for diagnosing and/or  
XX prognosticating disorders associated with an aberrant inflammatory  
XX response such as asthma.  
XX  
XX Disclosure; Fig 4; 290pp; English.  
XX  
XX The invention relates to an isolated human nucleic acid molecule  
XX comprising an allelic variant of a polymorphic region of a 5-lipoxygenase  
XX (5-LO) gene, where the allelic variant comprises one or more nucleotide  
XX selected from any of 3, 20 or 21 base pair sequences, given in the  
XX specification, or their complement. The compositions and methods of the

CC present invention are useful for diagnosing and/or prognosing disorders  
CC associated with an aberrant inflammatory response such as asthma,  
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,  
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic  
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory  
CC disease, atopic and contact dermatitis. The nucleic acid molecules can  
CC also be useful for identifying an individual amongst other individuals  
CC from the same species for use in forensic medicine and paternity testing.  
CC This polymucleotide sequence represents DNA relating to the human 5-  
CC lipoxygenase (5-LO) gene of the invention

XX  
SQ Sequence 168174 BP; 46808 A; 36442 C; 36942 G; 46474 T; 0 U; 1508 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 168174;  
Best Local Similarity 84.0%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
||||| ||||||| ||||| |||||  
Db 91147 AAAGAACTGCATCCTTAATGTTCTG 91171

RESULT 19  
ABT11114  
ID ABT11114 standard; DNA; 168273 BP.  
XX  
AC ABT11114;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 2.  
XX  
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;  
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;  
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;  
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;  
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;  
ds.

XX Homo sapiens.  
XX WO200262825-A2.  
XX 15-AUG-2002.  
XX 07-FEB-2002; 2002WO-US003546.  
XX  
XX 08-FEB-2001; 2001US-0267515P.  
XX 21-AUG-2001; 2001US-0314248P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX

XX Barnes G, Meyer J;  
XX WPI; 2002-627522/67.  
XX  
XX New isolated nucleic acid molecule with an allelic variant of a  
XX polymorphic region of an 5-LO gene, useful for diagnosing and/or  
XX prognosticating disorders associated with an aberrant inflammatory  
XX response such as asthma.  
XX  
XX Disclosure; Fig 2; 290pp; English.  
XX

XX The invention relates to an isolated human nucleic acid molecule  
XX comprising an allelic variant of a polymorphic region of a 5-lipoxygenase  
XX (5-LO) gene, where the allelic variant comprises one or more nucleotide  
XX selected from any of 3, 20 or 21 base pair sequences, given in the  
XX specification, or their complement. The compositions and methods of the  
XX present invention are useful for diagnosing and/or prognosing disorders  
XX associated with an aberrant inflammatory response such as asthma,  
XX bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,  
XX rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic  
XX rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory





```
SQ Sequence 521 BP; 180 A; 75 C; 122 G; 144 T; 0 U; 0 Other;
Query Match 72.8%; Score 18.2; DB 10; Length 521;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCCCAATGTTA 23
   ||||| ||||| ||||| ||||| |||||
Db 440 AAAAACTGCATCCCCCAATATAA 418

RESULT 24
ADE60727/c
ID ADE60727 standard; DNA; 521 BP.
XX
AC ADE60727;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene AV701053, SEQ ID NO 6639.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO ) GEN HOSPITAL CORP.
PA (FARB ) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
WPI; 2003-268312/26.
DR GENBANK; AV701053.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC
```

```
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIDO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 521 BP; 180 A; 75 C; 122 G; 144 T; 0 U; 0 Other;
Query Match 72.8%; Score 18.2; DB 10; Length 521;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCCCAATGTTA 23
   ||||| ||||| ||||| ||||| |||||
Db 440 AAAAACTGCATCCCCCAATATAA 418

RESULT 25
ADE60731/c
ID ADE60731 standard; DNA; 521 BP.
XX
AC ADE60731;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene AV701053, SEQ ID NO 6643.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO ) GEN HOSPITAL CORP.
PA (FARB ) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
WPI; 2003-268312/26.
DR GENBANK; AV701053.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC
```

CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 521 BP; 180 A; 75 C; 122 G; 144 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 10; Length 521;  
Best Local Similarity 87.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCATCCCAATGTTA 23  
|||||||  
Dd 440 AAAAAAAGTCATCCCAATATAA 418

RESULT 26

ACN43954  
ID ACN43954 standard; DNA; 95683 BP.

XX AC ACN43954;

XX 18-NOV-2004 (first entry)

XX Human genomic sequence hCG37570.

XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

XX Homo sapiens.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
PT comprises a nucleotide sequence.

XX Claim 1; SEQ ID NO 160; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published

XX SQ Sequence 95683 BP; 28002 A; 20734 C; 20466 G; 26481 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 11; Length 95683;  
Best Local Similarity 87.0%; Pred. No. 4.4e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAGTCATCCCAATGTTAT 24  
|||||||  
Dd 8156 AAAAAAGTCATCCCAATGTTAT 8178

RESULT 27

ADL13855  
ID ADL13855 standard; DNA; 152048 BP.

XX AC ADL13855;

XX 06-MAY-2004 (first entry)

XX Osteoarthritis-associated polymorphic nucleotide #387.

XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;  
KW joint space narrowing; osteophyte development; joint pain;  
KW osteoarthritis; SNP; single nucleotide polymorphism.

XX Homo sapiens.

XX WO2003054166-A2.

XX 03-JUL-2003.

XX 19-DEC-2002; 2002WO-US041225.

XX 20-DEC-2001; 2001US-0342603P.

XX (INCY-) INCYTE GENOMICS INC.

XX Jones KA, Schafer A;

XX WPI; 2003-559141/52.

XX Determining susceptibility of an individual to joint space narrowing,  
PT osteophyte development and/or joint pain comprises identifying whether  
PT the individual has at least one polymorphism in a polynucleotide encoding  
PT a protein.

XX Disclosure; SEQ ID NO 387; 297pp; English.

XX The invention relates to a method of determining susceptibility of an  
CC individual to joint space narrowing and/or osteophyte development and/or  
CC joint pain comprising identifying whether the individual has at least one  
CC polymorphism in a polynucleotide encoding at least one of the protein  
CC listed in the specification. The methods, composition and agent are  
CC useful for modulating the susceptibility of an individual to joint space  
CC narrowing and/or osteophyte development and/or joint pain that is  
CC associated with a disease, preferably osteoarthritis. The cell line and  
CC the non-human animal are useful for screening for an agent for diagnosing  
CC an individual having susceptibility to joint space narrowing and/or  
CC osteophyte development and/or joint pain. This sequence corresponds to  
CC the polynucleotide encoding a protein listed in the specification. (Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences).

XX SQ Sequence 152048 BP; 45197 A; 32444 C; 31900 G; 42505 T; 0 U; 2 Other;

Query Match 72.8%; Score 18.2; DB 10; Length 152048;  
Best Local Similarity 87.0%; Pred. No. 4.7e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGTCATCCCAATGTTA 23  
|||||||  
Dd 95029 AAAAAAGTCATCCCAATGTTA 95051

```

RESULT 28
ABQ31619
ID ABQ31619 standard; DNA; 578 BP.
XX
AC ABQ31619;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18210.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; db.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PP 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 578 BP; 213 A; 230 C; 77 G; 58 T; 0 U; 0 Other;
XX
Query Match 71.2%; Score 17.8; DB 6; Length 578;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGT 21
||||| |||||||
Db 176 AAAAAACGCATCCCAATTT 196

RESULT 29
ABQ31618/c
ID -ABQ31618 standard; DNA; 578 BP.

```

```

XX ABQ31618;
XX AC
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18209.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; db.
XX OS Homo sapiens.
XX PN WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP010074.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PP 05-SEP-2000; 2000DE-01044543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 578 BP; 58 A; 77 C; 230 G; 213 T; 0 U; 0 Other;
XX
Query Match 71.2%; Score 17.8; DB 6; Length 578;
Best Local Similarity 90.5%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGT 21
||||| |||||||
Db 403 AAAAAACGCATCCCAATTT 383

RESULT 30
ACA24290/c
ID ACA24290 standard; DNA; 1083 BP.
XX
AC ACA24290;
XX

```

DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #5947.  
DE Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Bacteroides fragilis.  
OS WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU20420.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 12160; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1083 BP; 311 A; 180 C; 274 G; 318 T; 0 U; 0 Other;  
  
Query Match 71.2%; Score 17.8; DB 8; Length 1083;  
Best Local Similarity 90.5%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAATGCATCCCCCAATGTT 22  
DB 633 AAAAATGCATCCCCCAATCTT 613  
  
RESULT 31  
AEB39175\_33/c  
Continuation (34 of 35) of AEB39175 from base 3300001 (L. pneumophila DNA SEQ ID NO 35077)  
WP Sequence split into 35 fragments LOCUS AEB39175 Accession Aeb39175  
WP Fragment Name Begin End  
WP AEB39175\_00 1 110000  
WP AEB39175\_01 100001 210000  
WP AEB39175\_02 200001 310000  
WP AEB39175\_03 300001 410000  
WP AEB39175\_04 400001 510000  
WP AEB39175\_05 500001 610000  
WP AEB39175\_06 600001 710000  
WP AEB39175\_07 700001 810000  
WP AEB39175\_08 800001 910000  
WP AEB39175\_09 900001 1010000  
WP AEB39175\_10 1000001 1110000  
WP AEB39175\_11 1100001 1210000  
WP AEB39175\_12 1200001 1310000  
WP AEB39175\_13 1300001 1410000  
WP AEB39175\_14 1400001 1510000  
WP AEB39175\_15 1500001 1610000  
WP AEB39175\_16 1600001 1710000  
WP AEB39175\_17 1700001 1810000  
WP AEB39175\_18 1800001 1910000  
WP AEB39175\_19 1900001 2010000  
WP AEB39175\_20 2000001 2110000  
WP AEB39175\_21 2100001 2210000  
WP AEB39175\_22 2200001 2310000  
WP AEB39175\_23 2300001 2410000  
WP AEB39175\_24 2400001 2510000  
WP AEB39175\_25 2500001 2610000  
WP AEB39175\_26 2600001 2710000  
WP AEB39175\_27 2700001 2810000  
WP AEB39175\_28 2800001 2910000  
WP AEB39175\_29 2900001 3010000  
WP AEB39175\_30 3000001 3110000  
WP AEB39175\_31 3100001 3210000  
WP AEB39175\_32 3200001 3310000  
WP AEB39175\_33 3300001 3410000  
WP AEB39175\_34 3400001 3503610  
  
Query Match 71.2%; Score 17.8; DB 14; Length 110000;  
Best Local Similarity 90.5%; Pred. No. 6.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 AACTGCATCCCCCAATGTTATG 25  
DB 12430 AACTGAATCCCCCAATGCTATG 12410  
  
RESULT 32  
AEB35716  
ID AEB35716 standard; DNA; 162147 BP.  
XX AEB35716;  
XX AC  
XX DT 08-SEP-2005 (first entry)  
XX DE L. pneumophila DNA SEQ ID NO 48.  
XX KW detection; infection; Antibacterial; Vaccine; ds; gene.  
XX OS Legionella pneumophila.  
XX FN WO2005049642-A2.  
XX PD 02-JUN-2005.



```
PF 23-SEP-2004; 2004WO-IB003578.
XX
PR 21-NOV-2003; 2003FR-00013687.
XX
XX (INSM ) INST PASTEUR.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (UPLY-) UNIV LYON 1 BERNARD CLAUDE.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;
PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;
PI Jarraud S;
XX
XX WPI; 2005-388305/40.
XX
XX New genome of Legionella pneumophila Paris strain and derived
PT polypeptides, useful for detection or identification of the strain and
PT for treatment and prevention of infections.
XX
XX Claim 1; SEQ ID NO 48; 660pp; English.
XX
XX The invention relates to an isolated or purified nucleotide sequences (I)
CC from Legionella pneumophila Paris strain. (II), and their related
CC sequences or fragments, are useful as primers and probes for detection
CC and amplification, including differentiation between the Paris and
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant
CC (hybrid) polypeptides (II). (II) are also useful for preparation of
CC specific antibodies (Ab), also used for detection/identification of
CC Legionella, and some (I), specifically those involved in synthesis of
CC surface proteins, are targets for identification of inhibitors. (II), or
CC vectors that contain (I), are useful as vaccines and immunogenic
CC compositions, for treatment and prevention of infections by L.
CC pneumophila. The present sequence represents a L. pneumophila DNA.
XX
XX Sequence 162147 BP; 49444 A; 30180 C; 33920 G; 49593 T; 0 U; 10 Other;
SQ
Query Match 71.2%; Score 17.8; DB 14; Length 162147;
Best Local Similarity 90.5%; Pred. No. 7.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 5 AACTGCATCCCCAATGTTATG 25
Db 141886 AACTGAATCCCAATGCTATG 141906
||||| ||||||| |||||
RESULT 33
ACH39621/c
ID ACH39621 standard; cDNA; 461 BP.
XX
XX ACH39621;
AC
XX 13-OCT-2003 (first entry)
DT
XX Human foetal brain cDNA #988.
DE
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
OS
XX US2003073623-A1.
PN
XX 17-APR-2003.
PD
XX 30-JUL-2001; 2001US-00918995.
PF
XX 30-JUL-2001; 2001US-00918995.
PR
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
```

```
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 26833; 44pp; English.
PS
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 461 BP; 160 A; 64 C; 96 G; 126 T; 0 U; 15 Other;
SQ
Query Match 70.4%; Score 17.6; DB 9; Length 461;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AAAAACTGCATCCCCAATGTTATG 25
Db 209 AAAAACTGCATCAACACTGTTAAG 186
||||| ||||||| |||||||
RESULT 34
AAH09944/c
ID AAH09944 standard; cDNA; 549 BP.
XX
XX AAH09944;
AC
XX 26-JUN-2001 (first entry)
DT
XX Human cDNA clone (3'-primer) SEQ ID NO:6779.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-00116126.
PF
XX 29-JUL-1999; 99JP-00248036.
XX
XX 27-AUG-1999; 99JP-00300253.
PR
XX 11-JAN-2000; 2000JP-00118776.
PR
XX 02-MAY-2000; 2000JP-00183767.
PR
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
```

PR	07-JUN-2000;	2000US-020346767
PR	28-JUN-2000;	2000US-020148861
PR	30-JUN-2000;	2000US-02151358P
PR	07-JUL-2000;	2000US-02166476P
PR	07-JUL-2000;	2000US-02168807P
PR	11-JUL-2000;	2000US-02174877P
PR	11-JUL-2000;	2000US-02174967P
PR	14-JUL-2000;	2000US-02182902P
PR	26-JUL-2000;	2000US-02209633P
PR	26-JUL-2000;	2000US-02209643P
PR	14-AUG-2000;	2000US-02245188P
PR	14-AUG-2000;	2000US-02245198P
PR	14-AUG-2000;	2000US-02252133P
PR	14-AUG-2000;	2000US-02252143P
PR	14-AUG-2000;	2000US-02252576P
PR	14-AUG-2000;	2000US-02252667P
PR	14-AUG-2000;	2000US-02252677P
PR	14-AUG-2000;	2000US-02257558P
PR	18-AUG-2000;	2000US-02262797P
PR	22-AUG-2000;	2000US-02266811P
PR	22-AUG-2000;	2000US-02268682P
PR	23-AUG-2000;	2000US-02271822P
PR	23-AUG-2000;	2000US-02270009P
PR	30-SEP-2000;	2000US-02289247P
PR	01-SEP-2000;	2000US-02282877P
PR	01-SEP-2000;	2000US-02293433P
PR	01-SEP-2000;	2000US-02293442P
PR	01-SEP-2000;	2000US-02293455P
PR	05-SEP-2000;	2000US-02295059P
PR	05-SEP-2000;	2000US-02295133P
PR	06-SEP-2000;	2000US-02304377P
PR	06-SEP-2000;	2000US-02304387P
PR	08-SEP-2000;	2000US-02312422P
PR	08-SEP-2000;	2000US-02312433P
PR	08-SEP-2000;	2000US-02312442P
PR	08-SEP-2000;	2000US-02314133P
PR	08-SEP-2000;	2000US-02314142P
PR	08-SEP-2000;	2000US-02320807P
PR	08-SEP-2000;	2000US-02320812P
PR	12-SEP-2000;	2000US-02319687P
PR	14-SEP-2000;	2000US-02330633P
PR	14-SEP-2000;	2000US-02330643P
PR	14-SEP-2000;	2000US-02330655P
PR	14-SEP-2000;	2000US-02342237P
PR	14-SEP-2000;	2000US-02342239P
PR	14-SEP-2000;	2000US-02342401P
PR	14-SEP-2000;	2000US-02342412P
PR	14-SEP-2000;	2000US-02342422P
PR	14-SEP-2000;	2000US-02342433P
PR	14-SEP-2000;	2000US-02342442P
PR	21-SEP-2000;	2000US-02349977P
PR	25-SEP-2000;	2000US-02349987P
PR	26-SEP-2000;	2000US-02354844P
PR	27-SEP-2000;	2000US-02358347P
PR	27-SEP-2000;	2000US-02358367P
PR	29-SEP-2000;	2000US-02363677P
PR	29-SEP-2000;	2000US-02363687P
PR	29-SEP-2000;	2000US-02363697P
PR	29-SEP-2000;	2000US-02363707P
PR	02-OCT-2000;	2000US-02368027P
PR	02-OCT-2000;	2000US-02370377P
PR	02-OCT-2000;	2000US-02370387P
PR	02-OCT-2000;	2000US-02370397P
PR	13-OCT-2000;	2000US-02370407P
PR	13-OCT-2000;	2000US-02399357P
PR	13-OCT-2000;	2000US-02399377P
PR	20-OCT-2000;	2000US-02409607P
PR	20-OCT-2000;	2000US-02412217P

PR 20-OCT-2000; 2000US-02411785P.  
PR 20-OCT-2000; 2000US-02411786P.  
PR 20-OCT-2000; 2000US-02411787P.  
PR 20-OCT-2000; 2000US-02411808P.  
PR 20-OCT-2000; 2000US-02411809P.  
PR 20-OCT-2000; 2000US-02411826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249257P.  
PR 17-NOV-2000; 2000US-0249259P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251855P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251859P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX PI  
XX PI  
XX PI  
XX WPT; 2001-476224/51.  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the respiratory system including respiratory cancers  
PT and also for testing and detection e.g. diagnosis.  
XX  
XX Disclosure; SED ID No 1132; 546pp; English.  
XX  
XX The present invention relates to the isolation of novel human respiratory  
CC antigens (AAU17695-AAU17975), and cDNA and genomic sequences encoding for  
CC these polypeptides. The sequences of the invention are useful for  
CC preventing, treating and/or prognosing disorders related to the  
CC respiratory system including throat disorders (e.g. vocal cord paralysis,

CC tonsillitis, and laryngitis), lung disorders e.g. pneumonia, allergic  
CC disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose  
CC disorders and cancers of the respiratory tissues e.g. lung cancer. The  
CC polynucleotide sequences of the invention are useful in gene therapy and  
CC antisense therapy. AAS28161-AAS28764 represent genomic sequences encoding  
CC for novel human respiratory antigens. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 596 BP; 201 A; 78 C; 92 G; 225 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 4; Length 596;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGTGCATCCCCCAATGTTATG 25  
DB 290 AAAAAGTGAATCCCCCAATGTTATG 267  
RESULT 36  
ADG41894/C  
ID ADG41894 standard; DNA; 596 BP.  
XX  
AC ADG41894;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Human respiratory system associated genomic DNA seq id 1132.  
XX  
KW antiinflammatory; antiallergic; antiasthmatic; cytostatic; gene therapy;  
KW human respiratory system antigen;  
KW human respiratory system associated polynucleotide;  
KW respiratory system disorder; throat disorder; vocal cord paralysis;  
KW tonsillitis; laryngitis; lung disorder; pneumonia; allergic disorder;  
KW asthma; eosinophilic pneumonia; pleurisy; cystic fibrosis; emphysema;  
KW histiocytosis; sarcoidosis; nose disorder; rhinitis; sinusitis; neoplasm;  
KW cancer; respiratory tissue cancer; throat cancer; lung cancer;  
KW cancer of the nose; gene therapy; chromosome identification; forensic;  
KW human respiratory system associated protein; ds; human.  
XX  
OS Homo sapiens.  
XX  
XX US2003215893-A1.  
XX  
XX 20-NOV-2003.  
XX  
XX 07-AUG-2002; 2002US-00212872.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 14-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-02255267P.  
PR 14-AUG-2000; 2000US-02255268P.  
PR 14-AUG-2000; 2000US-02255270P.  
PR 14-AUG-2000; 2000US-0225547P.  
PR 14-AUG-2000; 2000US-02255757P.  
PR 14-AUG-2000; 2000US-02255758P.  
PR 14-AUG-2000; 2000US-02255759P.  
PR 14-AUG-2000; 2000US-0225579P.  
PR 14-AUG-2000; 2000US-02256279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-02311242P.  
PR 08-SEP-2000; 2000US-02311243P.  
PR 08-SEP-2000; 2000US-02311244P.  
PR 08-SEP-2000; 2000US-0231113P.  
PR 08-SEP-2000; 2000US-0231141P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231368P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234423P.  
PR 21-SEP-2000; 2000US-0234477P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 28-SEP-2000; 2000US-0235935P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-024617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764860.  
PR 14-FEB-2002; 2002US-00074095.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
PI WPI; 2003-902033/82.  
XX  
XX Novel respiratory system antigen and polynucleotides encoding the polypeptides, useful for treating diagnosing, treating or preventing tonsillitis, pneumonia, asthma and cystic fibrosis, emphysema, throat cancer.  
PS Disclosure; SEQ ID NO 1132; 236pp; English.  
XX  
XX The invention describes an isolated polypeptide (I) comprising an amino acid sequence that is at least 90% identical to polypeptide fragment of any one of 299 respiratory system antigen sequences (PS) and having biological activity, polypeptide domain or epitope of PS, full-length protein of PS, or variant, allelic variant or species homolog of PS. (I) or a polynucleotide (II) encoding (I) is also useful for diagnosing a pathological condition or a susceptibility to a pathological condition in a subject which involves determining the presence or absence of mutation in (II) or determining the presence or amount of expression of (I) in a biological sample and diagnosing a pathological condition based on the result. The human respiratory system associated polynucleotides, the polypeptides encoded by them, and antibodies that immunospecifically bind these polypeptides are useful in diagnosis, treatment, prevention and/or prognosis of disorders of respiratory system such as throat disorders (e.g., vocal cord paralysis, tonsillitis, and laryngitis), lung disorders (e.g., pneumonia), allergic disorders, (e.g., asthma and eosinophilic pneumonia), pleurisy, cystic fibrosis, emphysema, histiocytosis,

CC sarcoidosis, nose disorders (rhinitis and sinusitis), neoplasms and/or  
CC cancers of respiratory tissues (e.g., throat cancer, lung cancer, and  
CC cancer of the nose). The polynucleotides are useful in gene therapy  
CC techniques, for chromosome identification, identifying individuals from

Query Match 70.4%; Score 17.6; DB 10; Length 596;

Best Local Similarity 83.3%; Pred. No. 4.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAGTCATCCCATGTTATG 25

Db 290 AAAAAGTCATCCCATGTTATG 267

RESULT 37

AD197668/c

ID AD197668 standard; DNA; 596 BP.

XX

XX

AC

AD197668;

XX

DT

04-NOV-2004

(first entry)

XX

Human respiratory system associated polypeptide-related DNA SeqID1132.

XX

respiratory system-related polypeptide; antiasthmatic; antibacterial;

KW antiinflammatory; cytostatic; antianaemic; antiallergic; gene therapy;

KW pneumonia; lung cancer; cystic fibrosis; asthma; sarcoidosis; rhinitis;

KW anaemia; leukaemia; inflammation; sinusitis;

XX chronic obstructive pulmonary disease; infectious disease; human; ds.

XX

OS

Homo sapiens.

XX

PN

US2003077704-A1.

XX

24-APR-2003.

XX

14-FEB-2002; 2002US-00074095.

XX

31-JAN-2000; 2000US-0179065P.

PR

04-FEB-2000; 2000US-0180628P.

PR

24-FEB-2000; 2000US-0184664P.

PR

02-MAR-2000; 2000US-0186350P.

PR

16-MAR-2000; 2000US-0189874P.

PR

17-MAR-2000; 2000US-0190076P.

PR

18-APR-2000; 2000US-0198123P.

PR

19-MAY-2000; 2000US-0205515P.

PR

07-JUN-2000; 2000US-0209467P.

PR

28-JUN-2000; 2000US-0214886P.

PR

30-JUN-2000; 2000US-0215135P.

PR

07-JUL-2000; 2000US-0216647P.

PR

07-JUL-2000; 2000US-0216880P.

PR

11-JUL-2000; 2000US-0217496P.

PR

14-JUL-2000; 2000US-0218290P.

PR

26-JUL-2000; 2000US-0220963P.

PR

26-JUL-2000; 2000US-0220964P.

PR

14-AUG-2000; 2000US-0224518P.

PR

14-AUG-2000; 2000US-0224519P.

PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232402P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0232404P.  
PR 21-SEP-2000; 2000US-0232423P.  
PR 21-SEP-2000; 2000US-0232424P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764860.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
FI WPI; 2003-765403/72.  
XX  
XX  
XX New human respiratory system-related polypeptide and genes, useful for  
PT treating, preventing or diagnosing e.g. pneumonia, lung cancer, cystic  
PT fibrosis, asthma, sarcoidosis, rhinitis, leukemia, inflammations or  
PT sinusitis.  
XX  
XX Disclosure; SEQ ID NO 1132; 202pp; English.  
XX  
XX This invention is related to a novel isolated polypeptide, which  
CC comprises a human respiratory system-related polypeptide, and the DNA  
CC sequence which encodes it. The invention may be useful for the  
CC development of compounds with an antiasthmatic, antibacterial,  
CC antiinflammatory, cytostatic, antianaemic or antiallergic activity. In  
CC addition, the sequences disclosed may be useful for gene therapy. The  
CC polypeptide or polynucleotide is useful for treating, preventing or  
CC ameliorating a medical condition, for example pneumonia, lung cancer,  
CC cystic fibrosis, asthma, sarcoidosis, rhinitis, anaemia, leukaemia,  
CC inflammations, sinusitis, chronic obstructive pulmonary disease or  
CC infectious diseases. The polypeptide or polynucleotide is also useful for  
CC diagnosing any of these diseases or a susceptibility to the disease. The  
CC present sequence is that of a human DNA sequence which is related to a  
CC human respiratory system associated gene of the invention.  
XX  
XX Sequence 596 BP; 201 A; 78 C; 92 G; 225 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 11; Length 596;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGTGCATCCCAATGTTATG 25  
|||||  
Db 290 AAAAAGTGAATCCCAATGTTATG 267  
|||||  
RESULT 38  
ABV56236/c  
ID ABV56236 standard; cDNA; 612 BP.  
XX

AC ABV56236;  
XX  
XX 17-SEP-2002 (first entry)  
XX  
XX Human prostate expression marker cDNA 56227.  
XX  
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US005171.  
XX  
XX 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
FI WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
XX Claim 1; Page 10854; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
XX Sequence 612 BP; 173 A; 134 C; 140 G; 165 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 5; Length 612;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGTGCATCCCAATGTTATG 25  
|||||  
Db 276 AAAATCCTCATCCCAATGTTATG 253  
|||||  
RESULT 39  
AAS28126  
ID AAS28126 standard; cDNA; 614 BP.  
XX  
XX AAS28126;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Novel cDNA encoding for human respiratory antigen #258.  
XX  
XX Human; respiratory antigen; respiratory disorder; throat disorder;  
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;  
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;

KW respiratory active; ss.  
XX Homo sapiens.  
OS WO200155448-A1.  
PN 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US001333.  
XX 31-JAN-2000; 2000US-0179065P.  
PF 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216847P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 03-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234937P.  
PR 25-SEP-2000; 2000US-0234938P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.



XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-476224/51.  
XX P-PSDB; AAU17942.  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the respiratory system including respiratory cancers  
PT and also for testing and detection e.g. diagnosis.  
XX Claim 4; SED ID No 268; 546pp; English.  
XX The present invention relates to the isolation of novel human respiratory  
CC antigens (AAU17685-AAU17975), and cDNA and genomic sequences encoding for  
CC these polypeptides. The sequences of the invention are useful for  
CC preventing, treating and/or prognosing disorders related to the  
CC respiratory system including throat disorders (e.g. vocal cord paralysis,  
CC tonsillitis, and laryngitis), lung disorders e.g. pneumonia, allergic  
CC disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose  
CC disorders and cancers of the respiratory tissues e.g. lung cancer. The  
CC polynucleotide sequences of the invention are useful in gene therapy and  
CC antisense therapy. AAS27869-AAS28159 encode for novel human respiratory  
CC antigens. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 614 BP; 231 A; 96 C; 80 G; 207 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 4; Length 614;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 AAAAAGTCATCCCAATGTATG 25  
||||| ||||| ||||| |||||  
Db 309 AAAAAGTCATCCCAATTTTCATG 332  
  
RESULT 40  
ADG41030  
ID ADG41030 standard; cDNA; 614 BP.  
XX  
XX ADG41030;  
XX  
XX 26-FEB-2004 (first entry)  
XX Human respiratory system associated protein cDNA seq id 268.  
XX  
XX antiinflammatory; antiallergic; antiasthmatic; cytostatic; gene therapy;  
KW respiratory system antigen;  
KW human respiratory system associated polynucleotide;  
KW respiratory system disorder; throat disorder; pneumonia; vocal cord paralysis;  
KW tonsillitis; laryngitis; lung disorder; pneumonia; allergic disorder;  
KW asthma; eosinophilic pneumonia; pleurisy; cystic fibrosis; emphysema;  
KW histiocytosis; sarcoidosis; nose disorder; rhinitis; sinusitis; neoplasm;  
KW cancer; respiratory tissue cancer; throat cancer; lung cancer;  
KW cancer of the nose; gene therapy; chromosome identification; forensic;  
KW human respiratory system associated protein; gene; ss; human.  
XX Homo sapiens.  
OS  
XX  
XX US2003215893-A1.  
XX  
XX 20-NOV-2003.  
XX  
XX 07-AUG-2002; 2002US-00212872.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 28-SEP-2000; 2000US-0235935P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-02339937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-02446474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249298P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764860.  
PR 14-FEB-2002; 2002US-00074095.  
XX (HUMA-) HUMAN GENOME SCI INC.  
FA Rosen CA, Ruben SM, Barash SC;  
XX P-PSDB; ADG41322.  
XX WPI; 2003-902033/82.  
XX P-PSDB; ADG41322.  
XX Novel respiratory system antigen and polynucleotides encoding the  
PT polypeptides, useful for treating diagnosing, treating or preventing  
PT tonsillitis, pneumonia, asthma and cystic fibrosis, emphysema, throat  
PT cancer.  
XX

PS Claim 1; SEQ ID NO 268; 236pp; English.  
XX The invention describes an isolated polypeptide (I) comprising an amino  
CC acid sequence that is at least 90% identical to polypeptide fragment of  
CC any one of 299 respiratory system antigen sequences (PS) and having  
CC biological activity, polypeptide domain or epitope of PS, full-length  
CC protein of PS, or variant, allelic variant or species homolog of PS. (I)  
CC or a polynucleotide (II) encoding (I) is also useful for diagnosing a  
CC pathological condition or a susceptibility to a pathological condition in  
CC a subject which involves determining the presence or absence of mutation  
CC in (II) or determining the presence or amount of expression of (I) in a  
CC biological sample and diagnosing a pathological condition based on the  
CC result. The human respiratory system associated polynucleotides, the  
CC polypeptides encoded by them, and antibodies that immunospecifically bind  
CC these polypeptides are useful in diagnosis, treatment, prevention and/or  
CC prognosis of disorders of respiratory system such as throat disorders  
CC (e.g., vocal cord paralysis, tonsillitis, and laryngitis), lung disorders  
CC (e.g., pneumonia), allergic disorders, (e.g., asthma and eosinophilic  
CC pneumonia), pleurisy, cystic fibrosis, emphysema, histiocytosis,  
CC sarcoidosis, nose disorders (rhinitis and sinusitis), neoplasms and/or  
CC cancers of respiratory tissues (e.g., throat cancer, lung cancer, and  
CC cancer of the nose). The polynucleotides are useful in gene therapy

Query Match 70.4%; Score 17.6; DB 10; Length 614;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGTCGATCCCAATGTTATG 25  
DB 309 AAAAAGTCGATCCCAATGTTATG 332  
|||||  
RESULT 41  
AD196804  
ID AD196804 standard; DNA; 614 BP.  
XX  
AC AD196804;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Human respiratory system associated gene SeqID268.  
XX  
KW respiratory system-related polypeptide; antiasthmatic; antibacterial;  
KW antiinflammatory; cytostatic; antianaemic; antiallergic; gene therapy;  
KW pneumonia; lung cancer; cystic fibrosis; asthma; sarcoidosis; rhinitis;  
KW anaemia; leukaemia; inflammation; sinusitis;  
KW chronic obstructive pulmonary disease; infectious disease; human; gene;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
PN US2003077704-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 14-FEB-2002; 2002US-00074095.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251989P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764860.

(HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-765403/72.

XX P-PSDB; ADI97096.

XX New human respiratory system-related polypeptide and genes, useful for treating, preventing or diagnosing e.g. pneumonia, lung cancer, cystic fibrosis, asthma, sarcoidosis, rhinitis, leukemia, inflammations or sinusitis.

XX Claim 1; SEQ ID NO 268; 202pp; English.

XX This invention is related to a novel isolated polypeptide, which comprises a human respiratory system-related polypeptide, and the DNA sequence which encodes it. The invention may be useful for the development of compounds with an antiasthmatic, antibacterial, antiinflammatory, cytostatic, antianaemic or antiallergic activity. In addition, the sequences disclosed may be useful for gene therapy. The polypeptide or polynucleotide is useful for treating, preventing or ameliorating a medical condition, for example pneumonia, lung cancer, cystic fibrosis, asthma, sarcoidosis, rhinitis, anaemia, leukaemia, inflammations, sinusitis, chronic obstructive pulmonary disease or

CC infectious diseases. The polypeptide or polynucleotide is also useful for  
CC diagnosing any of these diseases or a susceptibility to the disease. The  
CC present sequence is that of a respiratory system associated human gene of  
CC the invention.

XX Sequence 614 BP; 231 A; 96 C; 80 G; 207 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 11; Length 614;

Best Local Similarity 83.3%; Pred. No. 4.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAGTCATCCCAATGTTATG 25

||||| ||||| ||||| ||||| |||||

Db 309 AAAAAGTCATCCCAATGTTATG 332

RESULT 42

ACA53946

ID ACA53946 standard; DNA; 1086 BP.

XX

AC ACA53946;

XX

DT 19-JUN-2003 (first entry)

XX

DE Prokaryotic essential gene #35603.

XX

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX

OS Versinia pestis.

XX

PN WO20027183-A2.

XX

PD 03-OCT-2002.

XX

XX 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

XX (ELIT-) ELITRA PHARM INC.

PA

XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

XX WPI; 2003-029926/02.

DR P-PSDB; ABUS0076.

DR

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX

XX Claim 14; SEQ ID NO 41816; 1766pp; English.

PS

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1086 BP; 283 A; 233 C; 282 G; 288 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 8; Length 1086;

Best Local Similarity 83.3%; Pred. No. 4.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGTCATCCCAATGTTAT 24

||||| ||||| ||||| ||||| |||||

Db 333 AAAAAGTCATCCCAATGTTAT 356

RESULT 43

ACA43061

ID ACA43061 standard; DNA; 1506 BP.

XX

AC ACA43061;

XX

DT 19-JUN-2003 (first entry)

XX

XX Prokaryotic essential gene #24718.

DE

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX

XX Pasteurella multocida.

OS

XX WO20027183-A2.

PN

XX 03-OCT-2002.

XX

XX 21-MAR-2002; 2002WO-US009107.

XX

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

XX (ELIT-) ELITRA PHARM INC.

XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

XX WPI; 2003-029926/02.

DR P-PSDB; ABUS39191.

DR

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX

XX Claim 14; SEQ ID NO 30931; 1766pp; English.

PS

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1506 BP; 472 A; 316 C; 304 G; 414 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 8; Length 1506;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCCCAATGTTATG 25  
DB 263 AAAAACTGCATCGCCATGTTTG 286  
|||||

RESULT 44  
ACH94314/c  
ID ACH94314 standard; DNA; 1770 BP.

XX ACH94314;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polynucleotide seqid 109.

XX Recombinant expression vector; transcription regulatory element;  
XX Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX P-PSDB; ABO60763.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 109; 932pp; English.

CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention

XX Sequence 1770 BP; 305 A; 509 C; 569 G; 387 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 11; Length 1770;  
Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCCCAATGTTATG 25

DB 1338 AAAAACTTCATCCCCCAGGATATG 1315  
|||||

RESULT 45

ACH94414

ID ACH94414 standard; DNA; 1803 BP.

XX ACH94414;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polynucleotide seqid 209.

XX Recombinant expression vector; transcription regulatory element;  
XX Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX P-PSDB; ABO60863.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 209; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention

XX Sequence 1803 BP; 397 A; 573 C; 524 G; 309 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 11; Length 1803;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCCCAATGTTATG 25

DB 406 AAAAACTTCATCCCCCAGGATATG 429  
|||||

```
RESULT 46
ADR60058
ID ADR60058 standard; cDNA; 1951 BP.
XX
AC ADR60058;
AD
DT 02-DEC-2004 (first entry)
XX
DE Cotton cDNA sequence, SEQ ID 839.
XX
KW Cotton; ss; plant; cold tolerance; growth rate; cell cycle pathway;
KW drought tolerance; plant disease resistance; galactomannan; lignin;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW homologous recombination; extreme osmotic condition tolerance;
KW pathogen resistance; pest resistance; yield; photosynthesis; seed oil;
KW stress resistance.
XX
OS Gossypium hirsutum.
XX
PN US2004181830-A1.
XX
PD 16-SEP-2004.
XX
PF 29-JAN-2004; 2004US-00767795.
XX
PR 07-MAY-2001; 2001US-00849529.
PR 12-DEC-2001; 2001US-00021323.
XX
PA (KOVA/) KOVALIC D K.
PA (ZHOU/) ZHOU Y.
PA (CAOV/) CAO Y.
XX
PI Kovalic DK, Zhou Y, Cao Y;
XX
WIPI; 2004-667718/65.
XX
XX New recombinant nucleic acid molecules and polypeptides from Gossypium
PT hirsutum, useful for producing plants with improved biological
PT characteristics (e.g. improved plant cold or drought tolerance).
XX
XX Claim 1; SEQ ID NO 839; 14pp; English.
XX
CC The invention relates to a recombinant polynucleotide comprising any of
CC the 58798 Cotton plant cDNA sequences mentioned in the specification.
CC Also a recombinant polypeptide comprising any of the 58798 amino acid
CC sequences mentioned in the specification and producing a plant having an
CC improved property. Producing a plant having an improved property
CC comprises transforming a plant with a recombinant construct comprising a
CC promoter region functional in a plant cell operably joined to a
CC polynucleotide comprising a coding sequence for a polypeptide associated
CC with the property, and growing the transformed plant. The polypeptide is
CC useful for improving plant cold tolerance, manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, improving plant
CC drought tolerance, providing increased resistance to plant disease,
CC producing galactomannan (or lignin or plant growth regulators), improving
CC plant heat tolerance, improving plant tolerance to herbicides, increasing
CC the rate of homologous recombination in plants, improving plant tolerance
CC to extreme osmotic conditions or to pathogens or pests, improving yield
CC by modification of photosynthesis, modifying seed oil or protein yield
CC and/or content, improving yield by modification of carbohydrate, nitrogen
CC or phosphorus use and/or uptake, or improving yield by providing improved
CC plant growth and development under at least one stress condition. The
CC polynucleotide and polypeptide may also be used in recombinant DNA
CC constructs, in physical arrays of molecules, as plant breeding markers,
CC or in computer-based storage and analysis systems. The present sequence
CC is a Cotton plant cDNA of the invention. NOTE: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequences.html?docID=20040181830. However only 6585
CC polynucleotide sequences were available, the remaining 5213
CC polynucleotides and all 58798 protein sequences were not present.
XX
SQ .Sequence 1951 BP; 519 A; 417 C; 395 G; 620 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 13; Length 1951;
Best Local Similarity 83.3%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
DB 24 AAAAACTGCACCCCAATGTTTG 47

RESULT 47
ADB68941
ID ADB68941 standard; DNA; 2975 BP.
XX
AC ADB68941;
AD
DT 04-DEC-2003 (first entry)
XX
DE C. neoformans genomic DNA sequence SEQ ID NO:68.
XX
KW ds; gene; fungicide; gene therapy; infection.
XX
OS Cryptococcus neoformans.
XX
PN WO2003052076-A2.
XX
PD 26-JUN-2003.
XX
PF 17-DEC-2002; 2002WO-US040225.
XX
PR 17-DEC-2001; 2001US-0341261P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Zamudio C, Eroshkin AM;
XX
WIPI; 2003-533017/50.
XX
P-PSDB; ADB70024.
XX
XX New nucleic acid, useful for preparing a composition for treating an
PT infection caused by Cryptococcus neoformans.
XX
XX Claim 3; SEQ ID NO 68; 136pp; English.
XX
CC The invention relates to a novel purified or isolated Cryptococcus
CC neoformans nucleic acid molecule comprising a sequence encoding a
CC polypeptide comprising a sequence not given in the specification. A
CC polynucleotide of the invention has fungicide activity, and may have a
CC use in gene therapy. The nucleic acid is useful for preparing a
CC composition for treating an infection caused by Cryptococcus neoformans.
CC The present sequence represents a C. neoformans sequence of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2975 BP; 747 A; 739 C; 716 G; 772 T; 0 U; 1 Other;

Query Match 70.4%; Score 17.6; DB 10; Length 2975;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
DB 1974 AGAAGACGGCATCGACAATGTTAT 1997

RESULT 48
ADE25037/c
ID ADE25037 standard; cDNA; 3833 BP.
XX
AC ADE25037;
AD
DT 29-JAN-2004 (first entry)
```

```
XX DE Plant growth associated gene cDNA seq id 12.
XX KW plant growth; plant growth trait modulation; Brassicaceae; Arabidopsis;
XX KW Brassica; Zea; Oryza; Triticum; Hordeum; Lolium; Sorghum; Glycine;
KW Medicago; Helianthus; Lactuca; Beta; Vitis; Solanum; Lycopersicon;
KW Capsicum; Gossypium; Hevea; Linum; Prunus; Citrus; Populus; Pinus;
XX Quercus; gene; ss.
XX OS Magnoliophyta.
XX PN US2003188343-A1.
XX PD 02-OCT-2003.
XX PF 07-JAN-2003; 2003US-00338777.
XX PR 09-JAN-2002; 2002US-0347288P.
XX PA (LYNX-) LYNX THERAPEUTICS INC.
XX PI Bowen BA, Haudenschild CD, Buckler ES;
XX DR WPI; 2003-803305/75.
XX DR P-PSDB; ADE25067.
XX PT New isolated or recombinant polypeptide for use in modulating a plant
PT growth trait in a flowering plant e.g. in Arabidopsis, Brassica, Zea, or
PT Oryza.
XX PS Claim 1; SEQ ID NO 12; 81pp; English.
XX CC The invention describes an isolated or recombinant polypeptide (I)
CC comprising a sequence: (a) comprising 1 of 30 sequences (S1), as given in
CC the specification, or a conservative variant; (b) encoded by 1 of 30
CC sequences (S2), as given in the specification, or a conservative variant;
CC (c) encoded by a sequence that hybridises under stringent conditions to
CC S2; and (d) encoded by a sequence 70 % identical to S2. The expression or
CC activity of (I) is modulated to modulate a plant growth trait in a
CC flowering plant, of the family Brassicaceae, preferably in a plant that
CC is Arabidopsis, Brassica, Zea, Oryza, Triticum, Hordeum, Lolium, Sorghum,
CC Glycine, Medicago, Helianthus, Lactuca, Beta, Vitis, Solanum,
CC Lycopersicon, Capsicum, Gossypium, Hevea, Linum, Prunus, Citrus, Populus,
CC Pinus, or Quercus. A new method is used to detect genes for a plant
CC growth trait. This sequence represents a plant growth associated cDNA.
XX SQ Sequence 3833 BP; 1023 A; 705 C; 831 G; 1274 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 10; Length 3833;
Best Local Similarity 83.3%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAACTGCATCCCAATGTTAT 24
Db 1923 AAAAAATTCATCTCTCAATGTTAT 1900
RESULT 49
AAK68667/c
ID AAK68667 standard; DNA; 4963 BP.
XX AC AAK68667;
XX XX
XX 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23479.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
```





PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-02117487P.  
PR 11-JUL-2000; 2000US-02117496P.  
PR 14-JUL-2000; 2000US-02118290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0244647P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-451937/48.  
XX  
XX Isolated polypeptide for treating, preventing and/ or prognosing  
PT disorders related to the musculoskeletal system including musculoskeletal  
PT cancers and also for testing and detection e.g. diagnosis.  
XX  
XX Example 2; SEQ ID NO 2903; 781pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins  
CC (ABR03087-ABR04109) associated with the musculoskeletal system useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. The genes are isolated from a range of human tissues  
CC disclosed in the specification. The nucleic acids, proteins, antibodies  
CC and (ant)agonists are useful in the diagnosis, treatment and prevention  
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC

CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 4963 BP; 1480 A; 991 C; 1089 G; 1403 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 4; Length 4963;  
Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
||||| | | | | | | | | |  
Db 2660 AAAAACTGGATTCCCAAGATAT 2637

Search completed: February 3, 2006, 21:57:05  
Job time : 210.111 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-25

Perfect score: 25

Sequence: 1 aaaaactgcatcccaatgttatg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_hcc.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_est7.\*
- 9: gb\_gss1.\*
- 10: gb\_gss2.\*
- 11: gb\_gss3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	23.4	93.6	226	1	BB127535
2	23.4	93.6	274	1	AV259784
3	23.4	93.6	276	1	AV332113
C 4	23.4	93.6	340	6	CD776103
5	23.4	93.6	413	5	BY619222
C 6	23.4	93.6	419	1	AI847234
7	23.4	93.6	438	5	BY616767
8	23.4	93.6	449	2	BB750064
C 9	23.4	93.6	474	1	AI844638
10	23.4	93.6	648	1	BB130746
11	23.4	93.6	887	4	AK079373
12	22.4	89.6	175	1	BB083460
C 13	22.4	89.6	220	1	AI412710
14	22.4	89.6	279	2	BB347350
15	22.4	89.6	329	2	BB391724
C 16	22.4	89.6	369	2	BP404880
17	22.4	89.6	380	1	AA997380
C 18	21.8	87.2	512	3	BQ177061
19	21.8	87.2	647	2	BB629036
C 20	21.8	87.2	744	7	CO873358
21	21.8	87.2	764	7	CO878266
22	21	84.0	407	2	BG727345

C 23	20.2	80.8	602	5	BUS51162
24	20.2	80.8	698	10	CG818112
25	20.2	80.8	726	7	CO876233
26	20.2	80.8	830	6	CB206579
27	19.8	79.2	454	2	BES55733
28	19.8	79.2	561	3	BMS26844
C 29	19.8	79.2	783	3	BI970360
C 30	19.8	79.2	780	3	BI968779
31	19.4	77.6	407	7	CK028853
32	19.4	77.6	407	7	CNS10578
33	19.4	77.6	435	1	AI793511
34	19.4	77.6	494	1	AL730907
C 35	19.4	77.6	582	1	AL720246
36	19.4	77.6	595	1	AW116002
C 37	19.4	77.6	600	1	AL720381
38	19.4	77.6	720	1	AL723296
C 39	19.4	77.6	802	3	BI886461
40	19.4	77.6	807	10	BX188572
C 41	19.2	76.8	375	6	CD169592
42	19.2	76.8	403	8	H70808
43	19.2	76.8	580	9	AZ951719
44	19.2	76.8	649	11	CR252820
45	19.2	76.8	711	10	CZ789932
46	19.2	76.8	724	9	AQ543284
C 47	19.2	76.8	1052	6	CB561852
48	18.8	75.2	418	7	CO503266
C 49	18.8	75.2	431	7	CV041200
C 50	18.8	75.2	443	5	BU360170
51	18.8	75.2	502	2	BF288598
C 52	18.8	75.2	504	7	CV039923
C 53	18.8	75.2	570	5	BU326261
C 54	18.8	75.2	601	1	AW920398
C 55	18.8	75.2	620	5	BU403512
C 56	18.8	75.2	634	1	AI011772
C 57	18.8	75.2	640	9	CE241594
C 58	18.8	75.2	664	5	BU488926
C 59	18.8	75.2	669	5	BU406596
C 60	18.8	75.2	693	5	BU331120
C 61	18.8	75.2	727	5	BU308207
C 62	18.8	75.2	737	1	AJ395997
C 63	18.8	75.2	764	8	DR429774
C 64	18.8	75.2	783	5	BU401191
C 65	18.8	75.2	826	5	BU235584
C 66	18.8	75.2	860	5	BU118922
67	18.8	75.2	1080	8	DR149923
C 68	18.8	75.2	1112	10	CNS02KH6
C 69	18.6	74.4	182	2	BB482410
70	18.6	74.4	235	1	BB059223
71	18.6	74.4	265	1	AV325852
72	18.6	74.4	276	1	AW197966
C 73	18.6	74.4	304	3	BM260274
74	18.6	74.4	348	1	AJ774935
75	18.6	74.4	356	6	CF331098
C 76	18.6	74.4	390	6	CF794248
77	18.6	74.4	406	5	BY625606
78	18.6	74.4	453	6	CF792878
79	18.6	74.4	535	10	CW199101
C 80	18.6	74.4	535	10	CW199102
81	18.6	74.4	540	1	AW787162
82	18.6	74.4	542	2	BB692631
83	18.6	74.4	549	2	BI088166
C 84	18.6	74.4	549	8	T65864
C 85	18.6	74.4	551	2	BB695576
86	18.6	74.4	556	9	AZ879978
C 87	18.6	74.4	588	10	BX205218
88	18.6	74.4	618	6	CB858300
C 89	18.6	74.4	642	10	CW195394
C 90	18.6	74.4	652	1	AL729600
C 91	18.6	74.4	673	5	BU567394
C 92	18.6	74.4	680	10	CW889602
C 93	18.6	74.4	683	1	AV720574
94	18.6	74.4	723	2	BG107906
95	18.6	74.4	732	7	CK289165

BU551162	GM880019B
CG818112	SOYEK70TV
CO876233	BoVgen 01
CB206579	AGENCOURT
BE555733	EP53C03.Y
BM526844	BAL46411.
BI970360	GM830010A
BI968779	GM830006A
CK028853	AGENCOURT
CNS10578	AGENCOURT
AI793511	fc49b09.x
AL730907	AL730907
AL720246	AL720246
AW116002	fi106a09.x
AL720381	AL720381
AL723296	AL723296
BI886461	ZFG37-1-0
BX188572	Danio rer
CD169592	MMI-0024T
H70808	ys07a12.r1
AZ951719	2M0216F15
CR252820	Reverse s
CZ789932	OC_Ba015
AQ543284	RPCI-11-3
CB561852	AGENCOURT
CO503266	GGEZCB100
CV041200	4138919 B
BU360170	603477364
BF288598	EST453189
CV039923	4137361 B
BU326261	603492845
AW920398	EST351702
BU403512	604137361
AI011772	EST206223
CE241594	tiqr-988-
BU488926	604127456
BU406596	604136873
BU331120	603869085
BU308207	603536959
AJ395997	AJ395997
DR429774	naX2a09.
BU401191	603481805
BU235584	603791524
BU118922	603743589
DR149923	49232399
AL201507	Tetraodon
BB482410	BB482410
BB059223	BB059223
AV325852	AV325852
AW197966	x028b12.x
BM260274	baa18a09.
AJ774935	AJ774935
CF331098	NACL-07-
CF794248	889742 MA
BY625606	BY625606
CF792878	885870 MA
CW199101	104_624_1
CW199102	104_624_1
AW787162	120898 MA
BB692631	BB692631
BI088166	6028851203
T65864	YC24C03.r1
BB695576	BB695576
AZ879978	RPCI-23-1
BX205218	Danio rer
CB858300	HI06105W
CW195394	104_619_1
AL729600	AL729600
BU567394	AGENCOURT
CW889602	RPCI42_11
AV720574	AV720574
BG107906	602278180
CK289165	EST751887



QY 1 AAAAACTGCATCCCAATGTTATG 25  
 |||||  
 Db 150 AAAAACTGCATCCCAATGTTATG 174

RESULT 2  
 AV259784 274 bp mRNA linear EST 04-NOV-1999  
 LOCUS AV259784 RIKEN full-length enriched, adult male testis (DH10B) Mus  
 DEFINITION musculus cDNA clone 4930406D09 3' similar to X62840 R.rattus mRNA  
 for potassium channel protein (3145 bp), mRNA sequence.

ACCESSION AV259784  
 VERSION AV259784.1 GI:6247243  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,T., Wachihi,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al. 1999)  
 Unpublished (1999)

TITLE Contact: Yoshihide Hayashizaki  
 JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 COMMENT Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES  
 source Location/Qualifiers  
 1..274  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="4930406D09"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, adult male testis (DH10B)"  
 /notes="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

ORIGIN  
 Query Match 93.6%; Score 23.4; DB 1; Length 274;  
 Best Local Similarity 96.0%; Pred. No. 5.7;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAACTGCATCCCAATGTTATG 25  
 |||||  
 Db 216 AAAAACTGCATCCCAATGTTATG 240

RESULT 3  
 AV332113 276 bp mRNA linear EST 11-NOV-1999  
 LOCUS AV332113 RIKEN full-length enriched, adult male medulla oblongata  
 DEFINITION Mus musculus cDNA clone 6330530C15 3' similar to X62840 R.rattus  
 mRNA for potassium channel protein (3145 bp), mRNA sequence.

ACCESSION AV332113  
 VERSION AV332113.1 GI:6372165  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Wachihi,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al. 1999)  
 Unpublished (1999)

TITLE Contact: Yoshihide Hayashizaki  
 JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 COMMENT Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES  
 source Location/Qualifiers  
 1..274  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="4930406D09"  
 /sex="male"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, adult male testis (DH10B)"  
 /notes="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
 GAGAGAGAGAGATCCCAAGATCTTTTTTTTTTTTTTTN 3'], cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
 GAGAGAGAGATCTCGAGTTAATTAATTAATCCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."



source	sequence: 76-114, >AT_rich#Low_complexity (matched complement) Seq primer: M13 FORWARD POLYA=Yes.			
	FEATURES source			
1. .276	/organism="Mus musculus"			
	/mol_type="mRNA"			
/db_xref="taxon:10090"	/strain="C57BL/6J"			
	/clone="6330530C15"			
/sex="male"	/tissue_type="medulla oblongata"			
	/dev_stage="adult"			
/lab_host="DH10B"	/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"			
	/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',			
GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTT	GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTT			
	prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATCCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"			
ORIGIN				
Query Match 93.6%; Score 23.4; DB 1; Length 276; Best Local Similarity 96.0%; Pred. No. 5.7; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1 AAAAACTGCATCCCAATGTTATG 25			
Db	215 AAAAACTGCATCCCAATGTTATG 239			
RESULT 4	CD776103/c			
	LOCUS			
DEFINITION	CD776103 340 bp mRNA linear EST 02-JUL-2003			
	UI-M-AQ0-cjb-a-01-0-UI.s1 NIH_BMAP_MHI Mus musculus cDNA clone			
ACCESSION	CD776103			
	CD776103.1 GI:32434605			
KEYWORDS	EST.			
	Mus musculus (house mouse)			
SOURCE	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 340)			
	Bonaldo,M.F., Lennon,G. and Soares,M.B.			
AUTHORS	Normalization and subtraction: two approaches to facilitate gene discovery			
	Genome Res. 6 (9), 791-806 (1996)			
JOURNAL	8889548			
	Contact: Chin, H			
PUBMED	National Institute of Mental Health			
	6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD			
COMMENT	20892-9643, USA			
	Tel: 301 443 1706			
	Fax: 301 443 9890			
	Email: mEST@mail.nih.gov			
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa			
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa			
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa			
	Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).			
	The following repetitive elements were found in this cDNA			

FEATURES	Location/Qualifiers	
	source	
1. .340	/organism="Mus musculus"	
	/mol_type="mRNA"	
/db_xref="taxon:10090"	/strain="C57BL/6J"	
	/clone="UI-M-AQ0-cjb-a-01-0-UI"	
/dev_stage="27-32 days"	/lab_host="DH10B (Life Technologies)"	
	/clone_lib="NIH_BMAP_MHI"	
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MHI library is a non-normalized library constructed from mouse hippocampus. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.	TAG_TISSUE=hippocampus	
	TAG_LIB=UI-M-AQ0	
TAG_SEQ=TTGCA"	Query Match 93.6%; Score 23.4; DB 6; Length 340; Best Local Similarity 96.0%; Pred. No. 5.9; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
	ORIGIN	
QY	1 AAAAACTGCATCCCAATGTTATG 25	
Db	81 AAAAACTGCATCCCAATGTTATG 57	
RESULT 5	BY619222	
	LOCUS	
DEFINITION	BY619222 RIKEN full-length enriched, visual cortex Mus musculus	
	cDNA clone K330041H11 3', mRNA sequence.	
ACCESSION	BY619222	
	BY619222.1 GI:26954404	
KEYWORDS	EST.	
	Mus musculus (house mouse)	
SOURCE	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 413)	
	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrowsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,	

Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shimogawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 12466851  
 CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosewa,Wako-shi,Saitama 351-0198 Japan ) whose assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

#### FEATURES

Location/Qualifiers  
 1. .413  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="K330041H11"  
 /tissue types="visual cortex"  
 /clone\_lib="RIKEN full-length enriched, visual cortex"

#### ORIGIN

Query Match 93.6%; Score 23.4; DB 5; Length 413;  
 Best Local Similarity 96.0%; Pred. No. 6.2;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCCCAATGTTATG 25  
 |||||  
 Db 337 AAAAACTGCATGCCCAATGTTATG 361

#### RESULT 6

AI847234/c  
 LOCUS AI847234 419 bp mRNA linear EST 15-JUL-1999  
 DEFINITION UI-M-A11-afq-d-04-0-UI.s1 NIH-BMAP\_MBS\_N Mus musculus cDNA clone  
 UI-M-A11-afq-d-04-0-UI 3', mRNA sequence.  
 ACCSSION AI847234  
 VERSION AI847234.1 GI:5491140  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

#### ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 419)

#### REFERENCE

BONALDO,M.F., LENNON,G. and SOARES,M.B.  
 Normalization and subtraction: two approaches to facilitate gene discovery  
 Genome Res. 6 (9), 791-806 (1996)

#### JOURNAL

8889548  
 CONTACT: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mES@mail.nih.gov

#### COMMENT

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized brain stems library cDNA library Preparation: M.B. Soares lab Clone distribution: NIH-BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 180-217, >POLY\_A\$implete  
 Seq primer: M13 Forward  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1. .419  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-A11-afq-d-04-0-UI"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NIH-BMAP\_MBS\_N"  
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH-BMAP\_MBS\_N library is a normalized library constructed from mouse brain stems. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.  
 TAG\_TISSUE=brain-stems  
 TAG\_LIB=NIH-BMAP\_MBS\_N  
 TAG\_SEQ=TCATG"

#### ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 419;  
 Best Local Similarity 96.0%; Pred. No. 6.2;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCCCAATGTTATG 25  
 |||||  
 Db 185 AAAAACTGCATGCCCAATGTTATG 161

#### RESULT 7

BY616767  
 LOCUS BY616767 438 bp mRNA linear EST 15-DEC-2002  
 DEFINITION BY616767 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K330021F10 3', mRNA sequence.  
 ACCSSION BY616767  
 VERSION BY616767.1 GI:26951949  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)



FEATURES  
source

Location/Qualifiers  
1. .449  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="G130001113"  
/clone\_lib="RIKEN full-length enriched, pooled tissues, cerebellum, etc."  
/notes="pooled tissues ; (tissue\_type=cerebellum, dev\_stage=16 days neonate, sex=mixed), (tissue\_type=cerebellum, dev\_stage=0 day neonate, sex=mixed), (tissue\_type=hippocampus, dev\_stage=adult, sex=male), (tissue\_type=whole body, dev\_stage=9 days embryo, sex=mixed), (tissue\_type=lung, dev\_stage=13 days embryo, sex=mixed)"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 449;  
Best Local Similarity 96.0%; Pred. No. 6.3;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 57 AAAAACTGCATCCCAATGTTATG 81  
|||||

RESULT 9  
AI844638/c

LOCUS  
DEFINITION  
UI-M-ALI-ahr-b-07-0-UI.s1 NIH BMAP MCO N Mus musculus cDNA clone  
UI-M-ALI-ahr-b-07-0-UI 3', mRNA sequence.

ACCESSION  
AI844638  
VERSION  
AI844638.1 GI:5488544  
KEYWORDS  
EST.

SOURCE  
Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 474)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
normalized prefrontal cortex library cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made  
available by the means that is soon to be determined. When NIH  
determines the means for distribution of the BMAP cDNA clones, this  
record will be updated accordingly when that means is determined.  
The following repetitive elements were found in this cDNA sequence:  
169-206, >POLY A#Simple\_repeat  
Seq primer: M13 Forward  
POLYA=yes.

FEATURES  
source

Location/Qualifiers  
1. .474  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="UI-M-ALI-ahr-b-07-0-UI"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BMAP MCO\_N"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site1: NotI; Site2: Eco RI; The  
NIH BMAP MCO\_N library is a normalized library constructed  
from mouse cortex. The tag is a string of 5 nucleotides  
present between the NotI site and the oligo-dT track.  
The library was constructed as described by Bonaldo,  
Lennon and Soares, Genome Research 6: 791-806, 1996.  
Tissue provided by Ms. Annie Novakovich, Zivic-Miller  
Laboratories.  
TAG\_TISSUE=prefrontal-cortex  
TAG\_LIB=NIH BMAP\_MCO\_N  
TAG\_SEQ=CTCA"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 474;  
Best Local Similarity 96.0%; Pred. NO. 6.3;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 174 AAAAACTGCATCCCAATGTTATG 150  
|||||

## RESULT 10

BB130746

LOCUS  
DEFINITION

BB130746 RIKEN full-length enriched, 16 days neonate cerebellum Mus  
musculus cDNA clone 9630047A19 3', mRNA sequence.

ACCESSION  
BB130746

VERSION  
BB130746.2 GI:16266316

KEYWORDS  
EST.

SOURCE  
Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 648)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T.,  
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,  
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,  
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takanashi,F.,  
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jun 28, 2000 this sequence version replaced gi:8785136.

Contact: Yoshinhide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resesgsc.riken.jp, URL:http://genome.gsc.riken.jp/  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.  
and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

## FEATURES

source

1. .648  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="9630047A19"  
 /tissue\_type="cerebellum"  
 /dev\_stage="16 days neonate"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, 16 days neonate cerebellum"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTAAATTAATTAATTCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

## ORIGIN

Query Match 93.6%; Score 23.4; DB 1; Length 648;  
 Best Local Similarity 96.0%; Pred. No. 6.7;  
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25

DB 256 AAAAACTGCATCCCAATGTTATG 280

## RESULT 11

AK079373

LOCUS

AK079373 887 bp mRNA linear HTC 03-APR-2004  
 Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:9630047A19 product:unknown EST, full insert sequence.

ACCESSION

AK079373

VERSION

AK079373.1 GI:26098461

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

PUBMED

## REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kondo, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kondo, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

6 (bases 1 to 887)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.jp/>  
 URL: <http://fantom.gsc.riken.jp/>

## FEATURES

source

1. .887

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:9630047A19"

/db\_xref="taxon:10090"

/clone="9630047A19"

/tissue\_type="cerebellum"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="16 days neonate"

1. .887

/note="unknown EST (GB|BE750064, evidence: BLASTN, 99%, match=451)"

misc\_feature

## ORIGIN

Query Match 93.6%; Score 23.4; DB 4; Length 887;  
Best Local Similarity 96.0%; Pred. No. 7.2; 1; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
|||||  
495 AAAAACTGCATCCCAATGTTATG 519

## RESULT 12

BB083460  
LOCUS BB083460 RIKEN full-length enriched, adult male diencephalon Mus EST 28-JUN-2000  
DEFINITION musculus cDNA clone 9330182L24 3', mRNA sequence.

## ACCESSION

VERSION BB083460 GI:8648610

## KEYWORDS

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 175)

## AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Hayatsu, N.,  
Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,  
Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,  
Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,  
Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,  
Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A.,  
Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A.,  
Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and  
Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)

## TITLE

## JOURNAL

## COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsumai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES

## source

Location/Qualifiers  
1..175  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="9330182L24"  
/sex="male"  
/tissue\_type="diencephalon"  
/dev\_stages="adult"

## ORIGIN

Query Match 89.6%; Score 22.4; DB 1; Length 175;  
Best Local Similarity 95.8%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24

|||||  
113 AAAAACTGCATCCCAATGTTAT 136

## RESULT 13

## AI412710/C

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1..220

## /organism="Rattus sp."

## /mol\_type="mRNA"

## /db\_xref="taxon:10118"

## /clone="RBRDS47"

## /notes="Organ: brain; Vector: pT73Pac; Site\_1: EcoRI; Site\_2: NotI"

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 23; Conservative

## 0; Mismatches

## 1; Indels

## 0; Gaps

## 0;

/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, adult male  
diencephalon"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCGAGTTAAATAATATCCCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:  
BamHI"

Query Match 89.6%; Score 22.4; DB 1; Length 175;  
Best Local Similarity 95.8%; Pred. No. 15;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24

|||||

113 AAAAACTGCATCCCAATGTTAT 136

AI412710 220 bp mRNA linear EST 09-FEB-1999

EST241009 Normalized rat brain, Bento Soares Rattus sp. cDNA clone

RBRDS47 3' end, mRNA sequence.

AI412710 1 GI:4256214

EST.

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 220)

Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

Location/Qualifiers

1..220

/organism="Rattus sp."

/mol\_type="mRNA"

/db\_xref="taxon:10118"

/clone="RBRDS47"

/notes="Organ: brain; Vector: pT73Pac; Site\_1: EcoRI; Site\_2: NotI"

ORIGIN

Query Match 89.6%; Score 22.4; DB 1; Length 220;

Best Local Similarity 95.8%; Pred. No. 16;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0;









Genetics This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID=1773693 The following repetitive elements were found in this cDNA sequence: 75-114, >POLY\_A#simple\_repeat  
Seq primer: M13 Forward  
POLYA=No.

#### FEATURES

```

source
Location/Qualifiers
1..380
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CO-hq-f-10-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CO"
/notes="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-CO library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-CO) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-CO library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

```

#### ORIGIN

Query Match 89.6%; Score 22.4; DB 1; Length 380;  
Best Local Similarity 95.8%; Pred. No. 18;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 79 AAAAACTGCATCCCAATGTTATG 56

#### RESULT 18

B0177061/c  
LOCUS B0177061 512 bp mRNA linear EST 30-APR-2002  
DEFINITION UI-M-DJ2-bwa-b-13-0-UI.s1 NIH-BMAP\_DJ2 Mus musculus cDNA clone  
B0177061  
ACCESSION B0177061.1 GI:20352553  
VERSION  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 512)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
8889548  
JOURNAL PUBLISHED  
COMMENT Contact: Chin, H  
National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Tissue Procurement: Dr. Robin Davison  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
The following repetitive elements were found in this cDNA sequence: 170-207, >POLY\_A#simple\_repeat (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

#### FEATURES

```

source
Location/Qualifiers
1..512
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="UI-M-DJ2-bwa-b-13-0-UI"
/tissue_type="subfornical organ and postrema"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="NIH-BMAP_DJ2"
/notes="Organ: brain; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-M-DJ2 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is: GCTACATGAT, subfornical organ and area postrema.  
TAG_TISSUE=subfornical organ and postrema  
TAG_LIB=UI-M-DJ2  
TAG_SEQ=GCTACATGAT"

```

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 512;  
Best Local Similarity 92.0%; Pred. No. 36;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
|||||  
Db 175 AAAAACTGCATCCCAATGTTATG 151

#### RESULT 19

B0629036  
LOCUS B0629036 647 bp mRNA linear EST 26-OCT-2001  
DEFINITION B0629036 RIKEN full-length enriched, 16 days neonate cerebellum Mus musculus cDNA clone 9630047A19 5', mRNA sequence.  
B0629036  
ACCESSION B0629036  
VERSION B0629036.1 GI:16466302  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 647)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Harai,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

TITLE JOURNAL COMMENT	Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Arakawa,T., et al. 2001) Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., and Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.		RESULT 20 COB73358/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	COB73358 BovGen_01683 normal cattle brain Bos taurus cDNA clone RZPDp1056N1259Q 3', mRNA sequence. COB73358 COB73358.1 GI:51803198 EST. Bos taurus (cow) Bos taurus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos. 1 (bases 1 to 744) Hennig,S., Janitz,M., Herwig,R. and Williams,J. Generation, annotation, evolutionary analysis and database integration of 14969 cattle EST clusters Unpublished (2004) Contact: Hennig S Laboraty 123, dept. Lehrach Max-Planck-Institut fuer Molekulare Genetik Inhnstr.63-73, D-14195 Berlin, Germany Tel: +49 30 8413 1612 Fax: +49 30 8413 1380 Email: hennig@molgen.mpg.de The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and filters are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH (http://www.rzpd.de).
FEATURES source	1. .744 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="9630047A19" /tissue_type="cerebellum" /dev_stage="16 days neonate" /lab_host="DH10B" /clone_lib="RIKEN full-length enriched, 16 days neonate cerebellum" /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."		FEATURES Location/Qualifiers 1. .744 /organism="Bos taurus" /mol_type="mRNA" /db_xref="taxon:9913" /clone="RZPDp1056N1259Q" /sex="female" /tissue_type="brain tissue" /dev_stage="adult brain" /clone_lib="normal cattle brain" /note="Organ: brain; Vector: pSport1; Site 1: NotI; Site 2: SalI; Random primed and directionally cloned in pSport1 vector using NotI (5'-pGACTAGTCTTAGATCGGAGCGGCGGCC (T)15-3' and SalI 5'- TGCACCCACGCGTCCG-3' adapters (Gibco BRL))"	PCR Primers FORWARD: 5' CCCAGGCTTTACATTTAGTTCGGCTCG 3' (M13RSP) 5'-seq BACKWARD: 5' GCTATTACGCGAGTGGGAAAGGGGATGTG 3' (M13FSP) 3'-seq Seq primer: 5' GCTATTACGCGAGTGGGAAAGGGGATGTG 3' (M13FSP).
ORIGIN	Query Match 87.2%; Score 21.8; DB 2; Length 647; Best Local Similarity 92.0%; Pred. No. 38; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 491 AAAAACTGCTGCCCAATGTTATG 515		ORIGIN Query Match 87.2%; Score 21.8; DB 7; Length 744; Best Local Similarity 92.0%; Pred. No. 39; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 413 AAAAACTGCATCCCAATGTTATG 389	Query Match 87.2%; Score 21.8; DB 7; Length 744; Best Local Similarity 92.0%; Pred. No. 39; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 413 AAAAACTGCATCCCAATGTTATG 389
ORIGIN	Query Match 87.2%; Score 21.8; DB 2; Length 647; Best Local Similarity 92.0%; Pred. No. 38; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 491 AAAAACTGCTGCCCAATGTTATG 515		ORIGIN Query Match 87.2%; Score 21.8; DB 7; Length 744; Best Local Similarity 92.0%; Pred. No. 39; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 413 AAAAACTGCATCCCAATGTTATG 389	Query Match 87.2%; Score 21.8; DB 7; Length 744; Best Local Similarity 92.0%; Pred. No. 39; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTTATG 25                               DB 413 AAAAACTGCATCCCAATGTTATG 389

ORGANISM	Bos taurus	TITLE	WashU Zebrafish EST Project 1998
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	JOURNAL	Unpublished (1998)
AUTHORS	1 (bases 1 to 764)	COMMENT	Contact: Stephen L. Johnson Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: zbrafish@watson.wustl.edu
TITLE	Hennig, S., Janitz, M., Herwig, R. and Williams, J.		
JOURNAL	Generation, annotation, evolutionary analysis and database integration of 14969 cattle EST clusters		
COMMENT	Unpublished (2004)		
	Contact: Hennig S Laboraty 123, dept. Lehrach Max-Planck-Institut fuer Molekulare Genetik Inhnstr. 63-73, D-14195 Berlin, Germany Tel: +49 30 8413 1612 Fax: +49 30 8413 1380 Email: hennig@molgen.mpg.de		
	The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and filters are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH ( <a href="http://www.rzpd.de">http://www.rzpd.de</a> ).		
FEATURES	PCR Primers FORWARD: 5' CCCAGGCTTACACTTTATGCTCGGCTCG 3' (M13RSP) 5'-seq BACKWARD: 5' CTAATACCGACGTCGGCAAGGGGATGTG 3' (M13FSP) 3'-seq Seq primer: 5'-CCGTCGGAATTCGGGT-3' (M13RSP).	source	Seq primer: T7 from Gibco High quality sequence stop: 402. Location/Qualifiers 1..407 /organism="Danio rerio" /mol_type="mRNA" /db_xref="taxon:7955" /clone="IMAGE:4726428" /sex="mixed" /tissue_type="kidney pooled from 300 wild type adults" /lab_host="XL0LR" /clone_lib="zebrafish gridded kidney" /note="Organ: kidney; Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; Oligo dn cDNA library constructed from mRNA pooled from pooled kidney tissue from 300 adult zebrafish."
ORIGIN			
	Query Match 84.0%; Score 21; DB 2; Length 407; Best Local Similarity 100.0%; Pred. No. 83; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	2 AAAAACTGCATCCCAATGTT 22		
Db	198 AAAAACTGCATCCCAATGTT 218		
RESULT 23			
BU551162/c			
LOCUS	BU551162 602 bp mRNA linear EST 16-SEP-2002		
DEFINITION	GM880019B20D08 Gm-r1088 Glycine max cDNA clone Gm-r1088-7096 3', mRNA sequence.		
ACCESSION	BU551162		
VERSION	BU551162.1 GI:22934023		
KEYWORDS	EST.		
SOURCE	Glycine max (soybean)		
ORGANISM	Glycine max		
REFERENCE	1 (bases 1 to 602)		
AUTHORS	Vodkin, L., Shoemaker, R., Keim, P., Retzel, E., Khanna, A., Shealy, R., Clough, S., Thibaud-Nissen, F., Coryell, V., Erpeliding, J., Raph, C., Shoop, E., Stromvik, M., Schweitzer, P., Gong, G. and Liu, L.		
	A Functional Genomics Program for Soybean (NSF 9872565) (2002)		
TITLE	Unpublished (2002)		
JOURNAL	Other ESTs: BG726542 corresponding to Gm-cl067-4033 (5')		
COMMENT	Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565) Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics University of Illinois Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA Tel: (217) 244-6147 Fax: (217) 333-4582 Email: l-vodkin@uiuc.edu		

Insert Length: 602 Std Error: 0.00  
 Plate: GM80019B20 row: D column: 08  
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'  
 High quality sequence stop: 602.

## FEATURES

source

Location/Qualifiers  
 1..602

/organism="Glycine max"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3847"  
 /clone\_lib="Gm-r1088-7096"  
 /clone\_lib="Gm-r1088"  
 /notes="The library Gm-r1088 is a sequence-driven, reracked set of 9,216 low redundancy clones selected from cDNA libraries from various tissues and stages of development of soybean that consists of 2,706 cDNAs from germinating cotyledons (source library Gm-cl027); 1,355 cDNAs from immature seed coats (libraries Gm-cl019 and Gm-cl023); 917 cDNAs from tissue culture derived somatic embryos (source libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from germinating seedlings, shoot tips, or leaves exposed to various stresses (source libraries Gm-cl064, Gm-cl065, Gm-cl066, and Gm-cl067); and Gm-cl068); and 963 cDNAs from young leaves exposed to bacterial and fungal pathogens (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1088 and the cDNA clones of the reracked Gm-r1088 library were then sequenced at the 3' end. The unigene selection and 3' sequencing was funded by NSF Plant Genome project #9872565 (http://soybeanomics.cropsci.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://129.186.26.94/soybeanest.html. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://web.ahc.umn.edu/biodata/nfssoy/. Reracking and 3' sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 602;  
 Best Local Similarity 88.0%; Pred. No. 2.1e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTTATG 25

Db 499 ATAAAAATGCATCCCAATGTTAAG 475

## RESULT 24

CG818112

LOCUS

DEFINITION SOYEK70TV LargeInsertSoybeanGenLib Glycine max genomic clone

698 bp DNA linear GSS 18-NOV-2003

H65E06:MTP17L19, genomic survey sequence.

CG818112.1 GI:38273548

GSS.

KEYWORDS Glycine max (soybean)

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

## REFERENCE

AUTHORS

TITLE

## JOURNAL

COMMENT

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 1 (bases 1 to 698)  
 Shultz, J., Meksem, K., Shetty, J., Town, C. D., Koo, H., Potter, J., Wakefield, K., Zhang, H., Wu, C. and Lightfoot, D. A.  
 End sequencing of BACs comprising a provisional minimal tiling path from a fingerprint physical map of soybean (Glycine max) cultivar Forrest  
 Unpublished (2003)  
 Other GSSs: SOYEK70TH  
 Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
 The Center of Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University at Carbondale and Plant Genomics, The Institute for Genomic Research  
 Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415, USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 618 453 1797  
 Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL: http://bioinformatics.siu.edu)  
 Clones approximating a minimum tiling path were re-arrayed from the library master plates prior to sequencing.  
 For purposes of clone identification each clone name is a concatenation of the original clone location and its new location in the re-arrayed sequencing plates.  
 Seq primer: GTAATACGACTCACTATAGGG  
 Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..698

/organism="Glycine max"

/mol\_type="genomic DNA"

/cultivar="Forrest"

/db\_xref="taxon:3847"

/clone\_lib="H65E06:MTP17L19"

/note="Organ: Leaves; Vector: pCLD04541 (pBEL0BAC11 Scori clones); Site: BctXI; Soybean (Glycine max (L.) Merr.) cv. Forrest seeds were grown in greenhouse for fourteen days. Nuclei were isolated and embedded in agarose, restriction digested with Hind III BamHI or EcoRI, large size DNA fragments were ligated in vector V41 (pCLD04541) and electro transformed in DH10a cells. About 90,000 clones from BAC libraries were fingerprinted with HindIII and Hae III. Version 2 (automatic build) Contigs were built from 78,001 fingerprints. Contigs were manually examined to find the best non redundant tile path through the contigs representing about 13,000 clones. The clones were end sequenced."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 698;  
 Best Local Similarity 88.0%; Pred. No. 2.2e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTTATG 25

Db 394 ATAAAAATGCATCCCAATGTTAAG 418

## RESULT 25

CO876233

LOCUS

DEFINITION

BovGen 04558 normal cattle brain Bos taurus cDNA clone

R2PDp1056N1259Q 5', mRNA sequence.

CO876233.1 GI:51806145

EST.

KEYWORDS Bos taurus (cow)

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;

Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 726)

Hennig, S., Janitz, M., Herwig, R. and Williams, J.











cloned into the Not I and EcoRV sites of pExpress-1.  
Library was size-selected for >1 kb fragments and  
normalized. A non-normalized version of this library is  
also available (NIH ZGC 10). Library was constructed by  
Open Biosystems (Huntsville, AL)"

ORIGIN

Query Match      77.6%; Score 19.4; DB 7; Length 407;  
Best Local Similarity    95.2%; Pred.No. 4.7e+02;  
Matches    20; Conservative    0; Mismatches    1; Indels    0; Gaps    0;

QY    2   AAAAAGTCATCCCCCAATGTT 22  
         | |||||  
DB    209 AGAAACTGCATCCCCCAATGTT 229

RESULT 33  
AI793511          435 bp   mRNA   linear   EST 07-JUN-2001  
LOCUS   IC49B09.xl Zebrafish WashU MPIMG EST Danio rerio cDNA clone  
DEFINITION IMAGE:3724697 3', mRNA sequence.

ACCESSION AI793511  
VERSION AI793511.1 GI:5341227  
KEYWORDS EST  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 435)

REFERENCE  
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,  
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson.R.  
WashU Zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu

TITLE CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
JOURNAL Matthew Clark. DNA Sequencing by: Washington University Genome  
COMMENT Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ReisourcenzentrumPrimarDatenbank, Berlin, Germany (web address:  
www.rzpd.de)

Seq primer: T7 ET from Amersham  
High quality sequence stop: 419.  
Location/Qualifiers  
1. .435  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:3724697"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="XLI-blue MRF"  
/note="Vector: PSORTI; Site\_1: NotI; Site\_2: SalI; let  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
[5'-pgactagtcttagtcgcagcgcccccctttttttttttt13'];  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab: ICRF London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

FEATURES source

Location/Qualifiers  
1. .407  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7055972"  
/tissue\_type="whole body"  
/lab host="DH10B"  
/clone\_lib="NIH\_ZGC\_7"  
/note="Vector: pExpress1; Site\_1: NotI; Site\_2: EcoRV;  
Bulk tissue was collected from a whole adult individual  
from the Tuebingen strain. 1st strand cDNA was primed with  
a Not I - oligo(dT) primer, double-stranded cDNA was  
cloned into the Not I and EcoRV sites of pExpress-1.  
Library was size-selected for >1 kb fragments and  
normalized. A non-normalized version of this library is  
also available (NIH\_ZGC 10). Library was constructed by  
Open Biosystems (Huntsville, AL)"

ORIGIN

Query Match      77.6%; Score 19.4; DB 7; Length 407;  
Best Local Similarity    95.2%; Pred.No. 4.7e+02;  
Matches    20; Conservative    0; Mismatches    1; Indels    0; Gaps    0;

QY    2   AAAAAGTCATCCCCCAATGTT 22  
         | |||||  
DB    209 AGAACTGCATCCCCCAATGTT 229

RESULT 32  
CN510578          407 bp   mRNA   linear   EST 28-APR-2004  
LOCUS   AGENCOURT\_22490714 NIH\_ZGC\_7 Danio rerio cDNA clone IMAGE:7268330  
DEFINITION 5', mRNA sequence.

ACCESSION CN510578  
VERSION CN510578.1 GI:46823202  
KEYWORDS EST  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 407)

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rml0A07 Bethesda, MD 20892  
Email: cgabsr@mail.nih.gov

Tissue Procurement: Len Zon, Harvard  
CDNA Library Preparation: Open Biosystems  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM15247 row: h column: 24  
High quality sequence stop: 407.  
Location/Qualifiers  
1. .407  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7268330"  
/tissue\_type="whole body"  
/lab host="DH10B"  
/clone\_lib="NIH\_ZGC\_7"  
/note="Vector: pExpress1; Site\_1: NotI; Site\_2: EcoRV;  
Bulk tissue was collected from a whole adult individual  
from the Tuebingen strain. 1st strand cDNA was primed with  
a Not I - oligo(dT) primer, double-stranded cDNA was

FEATURES source

Location/Qualifiers  
1. .407  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7268330"  
/tissue\_type="whole body"  
/lab host="DH10B"



/db xref="taxon:7955"  
 /clone="IMAGE:2600344"  
 /sex="mixed (one male and one female, including unfertilized eggs)"  
 /dev\_stage="adult"

/lab\_host="DH10B (phage resistant)"  
 /clone\_lib="Sugano Kawakami zebrafish DRB"  
 /notes="Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCGTCTTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

## ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 595;  
 Best Local Similarity 95.2%; Pred. No. 5.1e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCCCAATGTT 22

Db 165 AGAAACTGCATCCCCCAATGTT 185

## RESULT 37

AL720381/c  
 LOCUS  
 DEFINITION AL720381 Danio rerio embryonic inner ear subtracted cDNA Danio rerio cDNA clone BN0AA045ZE03 5', mRNA sequence.

ACCESSION AL720381.1 GI:20184985

VERSION AL720381.1

KEYWORDS Danio rerio (zebrafish)

SOURCE Danio rerio

ORGANISM Danio rerio

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 600)

Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M.,

Hardelin,J.P., Weissenbach,J. and Petit,C.

A subtracted cDNA library from the zebrafish (Danio rerio)

embryonic inner ear

Unpublished (2002)

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..600

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="BN0AA045ZE03"

/tissue\_type="inner ear"

/dev\_stage="embryonic"

/clone\_lib="Danio rerio embryonic inner ear subtracted

cDNA"

/note="subtracted cDNA library"

## ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 600;  
 Best Local Similarity 95.2%; Pred. No. 5.1e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCCCAATGTT 22

Db 165 AGAAACTGCATCCCCCAATGTT 185

Db 410 AGAAACTGCATCCCCCAATGTT 390

## RESULT 38

AL723296

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..720

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="BN0AA064ZD10"

/tissue\_type="inner ear"

/dev\_stage="embryonic"

/clone\_lib="Danio rerio embryonic inner ear subtracted

cDNA"

/note="subtracted cDNA library"

Query Match 77.6%; Score 19.4; DB 1; Length 720;  
 Best Local Similarity 95.2%; Pred. No. 5.3e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCCCAATGTT 22

Db 98 AGAACTGCATCCCCCAATGTT 118

## RESULT 39

BI886461/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Hennig S

Laboraty 123, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1612

Fax: +49 30 8413 1380

AL723296 720 bp mRNA linear EST 18-APR-2002  
 rerio cDNA clone BN0AA064ZD10 5', mRNA sequence.

AL723296

AL723296.1 GI:20187900

EST

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 720)

Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M.,

Hardelin,J.P., Weissenbach,J. and Petit,C.

A subtracted cDNA library from the zebrafish (Danio rerio)

embryonic inner ear

Unpublished (2002)

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..720

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="BN0AA064ZD10"

/tissue\_type="inner ear"

/dev\_stage="embryonic"

/clone\_lib="Danio rerio embryonic inner ear subtracted

cDNA"

/note="subtracted cDNA library"

Query Match 77.6%; Score 19.4; DB 1; Length 720;  
 Best Local Similarity 95.2%; Pred. No. 5.3e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCCCAATGTT 22

Db 98 AGAACTGCATCCCCCAATGTT 118

BI886461 802 bp mRNA linear EST 12-OCT-2001  
 ZF637-1-000716 Zebrafish shield stage whole embryo cDNA library  
 MPMPGp637 Danio rerio cDNA clone MPMPGp637\_16N2;MPMPGp637N0216 5',  
 mRNA sequence.

BI886461

BI886461.1 GI:16093732

EST

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 802)

Clark,M., Aanstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.

EST sequencing of a zebrafish shield stage cDNA library normalised

by oligonucleotide fingerprinting

Unpublished (2001)

Contact: Hennig S

Laboraty 123, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1612

Fax: +49 30 8413 1380

Email: hennig@olgen.mpg.de  
 5' EST sequencing of clones from a zebrafish shield stage library,  
 normalised from 55,000 starting clones by oligonucleotide  
 fingerprinting  
 High quality sequence stop: 802.  
 Location/Qualifiers

# FEATURES

source  
 1..802  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="MPMGp637\_16N2;MPMGp637N0216"  
 /tissue type="whole embryo"  
 /dev stage="shield stage, 6 hrs post-fertilisation"  
 /lab\_host="E.coli XL1 blue MRF"  
 /clone\_lib="zebrafish shield stage whole embryo cDNA  
 library MPMGP637"  
 /note="Vector: pSort1; Site 1: NotI; Site 2: SalI;  
 oligo-dT-NotI primed, SalI adaptors, directionally cloned,  
 library normalised by oligonucleotide fingerprinting"

# ORIGIN

Query Match 77.6%; Score 19.4; DB 3; Length 802;  
 Best Local Similarity 95.2%; Pred. No. 5.4e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAAAACTGCATCCCAATGTT 22  
 Db 293 AGAAACTGCATCCCAATGTT 273

# RESULT 40

EX188572  
 LOCUS  
 DEFINITION  
 Danio rerio genomic clone DKEY-197G14, genomic survey sequence.  
 ACCESSION  
 EX188572  
 VERSION  
 EX188572.1 GI:28020458  
 KEYWORDS  
 GSS  
 SOURCE  
 Danio rerio (zebrafish)

# ORGANISM

Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 807)

# REFERENCE

AUTHORS  
 HUMPHRAY,S.J., HUCKLE,E. and DURHAM,J.L.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome  
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Unpublished

# COMMENT

This sequence was generated from the T7 end of BAC 197G14. 197G14  
 is part of the Daniokey BAC Library created by R. Plasterk and N.V.  
 Keygene. Further details:  
 http://www.sanger.ac.uk/Projects/D\_rerio/.

# FEATURES

source  
 1..807  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-197G14"  
 /tissue type="Testis"  
 /note="Vector pIndigoBAC-536"

# ORIGIN

Query Match 77.6%; Score 19.4; DB 10; Length 807;  
 Best Local Similarity 95.2%; Pred. No. 5.4e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTT 22

Db 534 AGAAACTGCATCCCAATGTT 554

# RESULT 41

CD169592

LOCUS  
 DEFINITION  
 MM1-0024T-M079-F09-U.G MM1-0024 Schistosoma mansoni cDNA clone  
 ACCESSION  
 CD169592  
 VERSION  
 CD169592.1 GI:34706196  
 KEYWORDS  
 EST.  
 SOURCE  
 Schistosoma mansoni  
 ORGANISM  
 Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.  
 REFERENCE  
 1 (bases 1 to 375)  
 AUTHORS  
 Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,  
 Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,  
 Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,  
 Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,  
 Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,  
 Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,  
 Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,  
 Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,  
 Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.  
 Transcriptional analysis of the acoelomate human parasite Schistosoma  
 mansoni

# TITLE

JOURNAL  
 Nat. Genet. 35 (2), 148-157 (2003)  
 PUBMED  
 12973350  
 COMMENT

Contact: Dr. Sergio Verjovski-Almeida  
 Departamento de Bioquímica  
 Instituto de Química - Universidade de São Paulo  
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,  
 Brasil  
 Tel: +55-11-3091-2173  
 Fax: +55-11-3091-2186  
 Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST  
 Genome Project. All sequences in the project were assembled and  
 annotated. This entry and all the assembled sequences can be seen  
 in the following URL <http://bioinfo.iq.usp.br/schisto/>  
 Plate: MM1-0024T-M079 row: 9 column: F.

# FEATURES

source  
 1..375  
 /organism="Schistosoma mansoni"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6183"  
 /clone="MM1-0024T-M079-F09.G"  
 /sex="male"  
 /dev stage="adult"  
 /lab\_host="Mesocricetus auratus"  
 /clone\_lib="MM1-0024"

# ORIGIN

Query Match 76.8%; Score 19.2; DB 6; Length 375;  
 Best Local Similarity 87.5%; Pred. No. 5.8e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24

Db 348 AAAAACTTCAGTCCCAATGTTAT 371

# RESULT 42

H70808/c

LOCUS

DEFINITION  
 Y807a12.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone  
 IMAGE:214078 5', mRNA sequence.

ACCESSION

H70808

VERSION

H70808.1 GI:1042624

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

REFERENCE

1 (bases 1 to 403)

AUTHORS

Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,



```

RESULT 46
AQ543284
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Query Match
Best Local Similarity
Matches
21; Conservative
0; Mismatches
3; Indels
0; Gaps
0;
Oy
1
539
AAAAAAAAATGCATCCCAATGTTAT
24
|||||
539
AAAAAAAACTGGACTCCCAATGTTAT
562
|||||
CB561852
1052 bp
mRNA
linear
EST 02-APR-2003
AGENCOURT
13322100
NICHDD_XGC_Tad1
Xenopus laevis
cDNA clone
CB561852
IMAGE:6880393
3', mRNA sequence.
CB561852
1
GI:29481382
EST.
Xenopus laevis
(African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1
(bases 1 to 1052)

```



AUTHORS TITLE JOURNAL COMMENT	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		/organism="Gallus gallus" /mol_type="mRNA" /db_xref="taxon:9031" /clone="GSEZCB1001D01" /tissue_type="breast muscle" /dev_stage="1 and 21 days old" /lab_host="DH5 alpha" /clone_lib="chicken breast muscle - CB1"	
	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs@mail.nih.gov">cgapbs@mail.nih.gov</a> Tissue Procurement: Drs. Donald Brown and Liguang Cai CDNA Library Preparation: CLONTECH CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCM3123 row: 1 column: 24 High quality sequence start: 169 High quality sequence stop: 516. Location/Qualifiers		/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This cDNA library was constructed with the SuperScript Plasmid System with Gateway technology kit (Invitrogen), following a manufacturer's protocols. Plasmid DNA was purified using a modified alkaline lysis method. Sequencing reactions were conducted using the DYEnamic Cycle Sequencing Kit for MegabACE (Amersham Biosciences) according to the manufacturer's recommendations. Clones were sequenced by the 5' end with T7 primer. Sequencing reactions were analyzed on MegabACE1000 DNA Sequencer (Amersham Biosciences). The quality and clustering of the ESTs were analyzed using the softwares Phred/Cap3. Only EST sequences with Phred quality greater than 20 and at least 150 bp were considered for clustering."	
FEATURES source	1. 1052 /organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:8355" /clone="IMAGE:6880393" /dev_stage="metamorphosis stage 53" /clone_lib="NICHD XGC Tadi" /note="Organ: Developing Tadpole; Vector: pDNR-LIB; Site 1: Sfi; Site 2: Sfi; 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCATATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.6 kb (range 0.9-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."		ORIGIN  Query Match 75.2%; Score 18.8; DB 7; Length 418; Best Local Similarity 90.9%; Pred. No. 9.1e+02; Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  QY 1 AAAAACTGCATCCCAATGTT 22       DB 233 AAAAAACTGCTTCCCAATTTT 212 	
	RESULT 49 CV041200/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM		CV041200 431 bp mRNA linear EST 01-OCT-2004 4138919 BARC_3GAL chicken mixed tissue Gallus gallus cDNA clone 3GAL 29009 5', mRNA sequence. CV041200 CV041200.1 GI:53560639 EST. Gallus gallus (chicken) Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 431) Evock-Clover,C.M., Ashwell,C.M., McMurtry,J.P., Lillehoj,H.S., Matukumalli,L.K. and Van Tassel,C.P. Characterization of expressed sequence tags generated from multiple chicken tissues Unpublished (2004) Contact: Christina M. Clover Growth Biology Laboratory Animal and Natural Resources Institute Bldg. 200 Rm215 BARC-East, Beltsville, MD 20705, USA Tel: 3015048224 Fax: 3015048623 Email: <a href="mailto:chrisc@anri.barc.usda.gov">chrisc@anri.barc.usda.gov</a> Single pass sequencing. Bases called and trimmed with phred 0.000925 using options -trim alt '' -trim fastavector identified by cross match using options -minmatch 12 -minscore 12 Plats: 29 row: 0 column: 09 Seq primer: CCAGTCACGACGCTGTGTAAACG High quality sequence stop: 431. Location/Qualifiers	
FEATURES source	1. 431 /organism="Gallus gallus" /mol_type="mRNA" /strain="Leghorn and broiler" /db_xref="taxon:9031" /clone="3GAL 29009" /lab_host="DH5alpha"		REFERENCE AUTHORS TITLE JOURNAL COMMENT	
	CV041200/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM		CV041200 418 bp mRNA linear EST 13-JUL-2004 GSEZCB1001D01.g chicken breast muscle - CB1 Gallus gallus cDNA clone GSEZCB1001D01, mRNA sequence. CV0503266 CV0503266.1 GI:50273452 EST. Gallus gallus (chicken) Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 418) Alves,H.J., Jorge,E.C., Marchesin,M.L., Monteiro-Vitorello,C.B., Patricio,M., Leduc,M.C. and Coutinho,L.L. Discovery of new genes expressed in the chicken breast muscle Unpublished (2004) Contact: Helena J. Alves Laboratory of Animal Biotechnology, Dep. of Animal Production ESALQ - University of Sao Paulo Av. Pádua Dias, 11, Piracicaba, SP, 13418-900, Brazil Tel: 55 19 3429 4434 Fax: 55 19 3429 4285 Email: <a href="mailto:hjalves@esalq.usp.br">hjalves@esalq.usp.br</a> and <a href="mailto:llcoutin@esalq.usp.br">llcoutin@esalq.usp.br</a> PCR Primers BACKWARD: T7. Location/Qualifiers	
FEATURES source	1. 418		REFERENCE AUTHORS TITLE JOURNAL COMMENT	
	CV0503266/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM		CV0503266 418 bp mRNA linear EST 13-JUL-2004 GSEZCB1001D01.g chicken breast muscle - CB1 Gallus gallus cDNA clone GSEZCB1001D01, mRNA sequence. CV0503266 CV0503266.1 GI:50273452 EST. Gallus gallus (chicken) Gallus gallus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 418) Alves,H.J., Jorge,E.C., Marchesin,M.L., Monteiro-Vitorello,C.B., Patricio,M., Leduc,M.C. and Coutinho,L.L. Discovery of new genes expressed in the chicken breast muscle Unpublished (2004) Contact: Helena J. Alves Laboratory of Animal Biotechnology, Dep. of Animal Production ESALQ - University of Sao Paulo Av. Pádua Dias, 11, Piracicaba, SP, 13418-900, Brazil Tel: 55 19 3429 4434 Fax: 55 19 3429 4285 Email: <a href="mailto:hjalves@esalq.usp.br">hjalves@esalq.usp.br</a> and <a href="mailto:llcoutin@esalq.usp.br">llcoutin@esalq.usp.br</a> PCR Primers BACKWARD: T7. Location/Qualifiers	

/clone lib="BARC 3GAL chicken mixed tissue"  
/note="Vector: pBluescript, SK+, Stratagene; Site 1: NotI;  
Site 2: EcoRI; Normalized library of pooled RNA isolated  
from whole brain, ultimobranchial gland, parathyroid  
gland, cecal tonsil and primordial germ cells Multiple"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 7; Length 431;  
Best Local Similarity 90.9%; Pred. No. 9.1e+02; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTT 22  
|||||  
Db 206 AAAAACTGCTTCCCAATTT 185

## RESULT 50

BU360170/c

LOCUS

BU360170 443 bp mRNA linear EST 28-NOV-2002  
603477364F1 CSEQCHN71 Gallus gallus cDNA clone ChEST363015 5', mRNA

DEFINITION

sequence.

ACCESSION

BU360170

VERSION

BU360170.1

KEYWORDS

EST.

SOURCE

Gallus gallus (chicken)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 443)

Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,

Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1. .443

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hisex"

/db\_xref="taxon:9031"

/clone="ChEST363015"

/dev\_stage="36"

/lab\_host="DH10B"

/clone\_lib="CSEQCHN71"

/note="Organ: hearts; Vector: pBluescript II KS(+);

Site 1: EcoRI; Site 2: NotI; This normalized library was

constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methylated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was blunted, ligated to NotI adapters, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pBluescript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 443;  
Best Local Similarity 90.9%; Pred. No. 9.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTT 22  
|||||  
Db 177 AAAAACTGCTTCCCAATTT 156

Search completed: February 3, 2006, 22:02:31  
Job time : 2961.67 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-25

Perfect score: 25

Sequence: 1 aaaaactgcattcccaatgttatg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
  - 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/ina/backfilese1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.8	87.2	270	3	US-09-016-434-1010
2	21.8	87.2	460	3	US-09-621-376-1487
3	20.4	81.6	1917	3	US-09-710-279-1467
4	20.4	81.6	1953	3	US-09-134-001C-1607
5	20.4	81.6	3276	3	US-09-710-279-4240
6	19.2	76.8	601	3	US-09-949-016-179347
7	19.2	76.8	233693	3	US-09-949-016-12386
8	19.2	76.8	233694	3	US-09-949-016-12386
9	18.6	74.4	168174	3	US-10-071-411A-63
10	18.6	74.4	168273	3	US-10-071-411A-2
11	17.8	71.2	300598	3	US-09-949-016-11868
12	17.8	71.2	302604	3	US-09-949-016-14588
13	17.8	71.2	302604	3	US-09-949-016-14588
14	17.8	71.2	308362	3	US-09-949-016-17119
15	17.6	70.4	601	3	US-09-949-016-62866
16	17.6	70.4	1770	3	US-09-489-039A-109
17	17.6	70.4	1803	3	US-09-489-039A-209
18	17.6	70.4	12563	3	US-09-949-016-4982
19	17.6	70.4	103792	3	US-09-949-016-13553
20	17.6	70.4	103993	3	US-09-949-016-12317
21	17.6	70.4	138693	3	US-09-949-016-16724
22	17.6	70.4	285986	3	US-09-949-016-12287
23	17.6	70.4	288031	3	US-09-949-016-14864
24	17.4	69.6	2179	3	US-09-620-312D-78
					Sequence 1010, Ap
					Sequence 14487, A
					Sequence 1467, Ap
					Sequence 1607, Ap
					Sequence 4240, Ap
					Sequence 179347, A
					Sequence 12386, A
					Sequence 16915, A
					Sequence 63, Appl
					Sequence 2, Appl
					Sequence 11868, A
					Sequence 14588, A
					Sequence 14589, A
					Sequence 17119, A
					Sequence 62866, A
					Sequence 109, App
					Sequence 209, App
					Sequence 4982, Ap
					Sequence 13553, A
					Sequence 12317, A
					Sequence 16724, A
					Sequence 12287, A
					Sequence 14864, A
					Sequence 78, Appl

25	17.2	68.8	881	3	US-09-270-767-11666	Sequence 11666, A
26	17.2	68.8	3198	3	US-09-345-236B-83	Sequence 83, Appl
C 27	17.2	68.8	3198	3	US-09-345-236B-87	Sequence 87, Appl
C 28	17.2	68.8	3198	3	US-09-345-236B-90	Sequence 90, Appl
C 29	17.2	68.8	3198	3	US-09-345-236B-92	Sequence 92, Appl
30	17.2	68.8	3918	3	US-08-836-323-1	Sequence 1, Appl
31	17.2	68.8	22605	3	US-09-949-016-17001	Sequence 17001, A
32	17.2	68.8	40655	3	US-09-949-016-12032	Sequence 12032, A
33	17.2	68.8	40655	3	US-09-949-016-15919	Sequence 15919, A
34	17.2	68.8	41696	3	US-09-949-016-12444	Sequence 12444, A
35	17.2	68.8	58782	3	US-09-949-016-16851	Sequence 16851, A
C 36	17.2	68.8	94142	3	US-09-949-016-18553	Sequence 16553, A
37	17.2	68.8	157822	3	US-09-949-016-16723	Sequence 16723, A
C 38	17	68.0	134	3	US-09-513-999C-16990	Sequence 16990, A
C 39	17	68.0	171	3	US-09-513-999C-22791	Sequence 22791, A
C 40	17	68.0	173	3	US-09-270-767-6867	Sequence 6867, Ap
C 41	17	68.0	173	3	US-09-270-767-22149	Sequence 22149, A
C 42	17	68.0	243	3	US-09-134-000C-150	Sequence 150, App
C 43	17	68.0	430	3	US-09-270-767-4566	Sequence 4566, Ap
44	17	68.0	430	3	US-09-270-767-19848	Sequence 19848, A
C 45	17	68.0	601	3	US-09-949-016-131417	Sequence 131417, A
46	17	68.0	654	3	US-09-543-681A-3477	Sequence 3477, Ap
47	17	68.0	918	3	US-09-328-352-2668	Sequence 2668, Ap
48	17	68.0	20229	3	US-09-949-016-18649	Sequence 16649, A
49	17	68.0	76810	3	US-09-949-016-12528	Sequence 12528, A
C 50	17	68.0	107820	3	US-09-792-616-1	Sequence 1, Appl
C 51	17	68.0	124700	3	US-09-949-016-11817	Sequence 11817, A
C 52	17	68.0	124701	3	US-09-949-016-15439	Sequence 15439, A
C 53	17	68.0	135171	3	US-09-949-016-15617	Sequence 15617, A
C 54	17	68.0	150409	3	US-09-949-016-12290	Sequence 12290, A
C 55	17	68.0	150409	3	US-09-949-016-12938	Sequence 12938, A
56	17	68.0	786431	3	US-09-751-389-3	Sequence 3, Appl
C 57	16.8	67.2	245	3	US-09-513-999C-8833	Sequence 8833, Ap
C 58	16.8	67.2	601	3	US-09-949-016-83336	Sequence 8336, A
59	16.8	67.2	1200	3	US-09-602-472A-11	Sequence 11, Appl
60	16.8	67.2	3456	3	US-09-602-472A-12	Sequence 12, Appl
61	16.8	67.2	143550	3	US-09-949-016-14143	Sequence 14143, A
C 62	16.8	67.2	168575	3	US-09-426-290-1	Sequence 1, Appl
C 63	16.8	67.2	392000	3	US-10-027-983-11	Sequence 11, Appl
C 64	16.6	66.4	380	3	US-09-347-798-5	Sequence 5, Appl
65	16.6	66.4	426	3	US-09-270-767-5056	Sequence 5056, Ap
66	16.6	66.4	426	3	US-09-270-767-20338	Sequence 20338, A
C 67	16.6	66.4	474	3	US-09-489-039A-7044	Sequence 7044, Ap
C 68	16.6	66.4	601	3	US-09-949-016-55241	Sequence 55241, A
C 69	16.6	66.4	601	3	US-09-949-016-74666	Sequence 74666, A
C 70	16.6	66.4	601	3	US-09-949-016-74865	Sequence 74865, A
C 71	16.6	66.4	601	3	US-09-949-016-75064	Sequence 75064, A
C 72	16.6	66.4	601	3	US-09-949-016-75263	Sequence 75263, A
C 73	16.6	66.4	601	3	US-09-949-016-75462	Sequence 75462, A
C 74	16.6	66.4	601	3	US-09-949-016-77347	Sequence 77347, A
C 75	16.6	66.4	601	3	US-09-949-016-77348	Sequence 77348, A
C 76	16.6	66.4	601	3	US-09-949-016-106020	Sequence 106020, A
C 77	16.6	66.4	601	3	US-09-949-016-106219	Sequence 106219, A
C 78	16.6	66.4	601	3	US-09-949-016-106418	Sequence 106418, A
C 79	16.6	66.4	601	3	US-09-949-016-106617	Sequence 106617, A
C 80	16.6	66.4	601	3	US-09-949-016-106816	Sequence 106816, A
C 81	16.6	66.4	601	3	US-09-949-016-128729	Sequence 128729, A
82	16.6	66.4	601	3	US-09-949-016-153399	Sequence 153399, A
83	16.6	66.4	601	3	US-09-949-016-153400	Sequence 153400, A
84	16.6	66.4	601	3	US-09-949-016-153401	Sequence 153401, A
85	16.6	66.4	601	3	US-09-949-016-153402	Sequence 153402, A
C 86	16.6	66.4	700	3	US-09-504-358-13	Sequence 1259, Ap
87	16.6	66.4	771	3	US-09-504-358-13	Sequence 13, Appl
88	16.6	66.4	771	3	US-09-504-358-13	Sequence 13, Appl
89	16.6	66.4	771	3	US-10-230-562-13	Sequence 9, Appl
C 90	16.6	66.4	1342	3	US-09-500-569-9	Sequence 9, Appl
C 91	16.6	66.4	1342	3	US-09-971-823B-94	Sequence 94, Appl
C 92	16.6	66.4	2161	3	US-09-509-712B-99	Sequence 7111, Ap
93	16.6	66.4	2439	2	US-09-489-039A-7111	Sequence 1, Appl
94	16.6	66.4	2814	2	US-07-781-034-1	Sequence 1, Appl
95	16.6	66.4	2814	6	PCT-US92-08328-1	Sequence 1, Appl
96	16.6	66.4	9848	3	US-09-385-222A-3	Sequence 3, Appl
97	16.6	66.4	10629	3	US-09-504-358-15	Sequence 15, Appl

98 16.6 66.4 10629 3 US-09-954-314-15 Sequence 15, Appl  
99 16.6 66.4 10629 3 US-10-230-562-15 Sequence 15, Appl  
100 16.6 66.4 11543 3 US-09-949-016-12071 Sequence 12071, A  
101 16.6 66.4 11543 3 US-09-949-016-12673 Sequence 12673, A  
102 16.6 66.4 11544 3 US-09-949-016-16172 Sequence 16172, A  
103 16.6 66.4 11544 3 US-09-949-016-16172 Sequence 16172, A  
104 16.6 66.4 11544 3 US-09-949-016-16174 Sequence 16174, A  
105 16.6 66.4 11544 3 US-09-949-016-16175 Sequence 16175, A  
106 16.6 66.4 12961 3 US-09-949-016-12902 Sequence 12902, A  
107 16.6 66.4 61399 3 US-09-949-016-14386 Sequence 14386, A  
108 16.6 66.4 101349 3 US-09-949-016-17433 Sequence 17433, A  
109 16.6 66.4 105733 3 US-09-949-016-13080 Sequence 13080, A  
110 16.6 66.4 113538 3 US-09-949-016-16329 Sequence 16329, A  
111 16.6 66.4 124110 3 US-09-949-016-13353 Sequence 13353, A  
112 16.6 66.4 130563 3 US-09-949-016-12273 Sequence 12273, A  
113 16.6 66.4 131379 3 US-09-949-016-16050 Sequence 16050, A  
114 16.6 66.4 154626 3 US-09-949-016-14000 Sequence 14000, A  
115 16.6 66.4 194790 3 US-09-949-016-15393 Sequence 15393, A  
116 16.6 66.4 228851 3 US-09-949-016-13781 Sequence 13781, A  
117 16.6 66.4 278866 3 US-09-949-016-13922 Sequence 13922, A  
118 16.6 66.4 278866 3 US-09-949-016-13923 Sequence 13923, A  
119 16.6 66.4 278866 3 US-09-949-016-13924 Sequence 13924, A  
120 16.6 66.4 278866 3 US-09-949-016-13925 Sequence 13925, A  
121 16.6 66.4 278866 3 US-09-949-016-13926 Sequence 13926, A  
122 16.6 66.4 278866 3 US-09-949-016-14699 Sequence 14699, A  
123 16.6 66.4 278866 3 US-09-949-016-14700 Sequence 14700, A  
124 16.6 66.4 278866 3 US-09-949-016-14701 Sequence 14701, A  
125 16.6 66.4 278866 3 US-09-949-016-14702 Sequence 14702, A  
126 16.6 66.4 278866 3 US-09-949-016-14703 Sequence 14703, A  
127 16.6 66.4 580073 3 US-08-545-528D-1 Sequence 1, Appli  
128 16.6 66.4 1664976 3 US-08-916-421B-1 Sequence 1, Appli  
129 16.6 66.4 1664976 3 US-08-916-421B-1 Sequence 1, Appli  
130 16.6 66.4 1664976 3 US-09-692-570-1 Sequence 1, Appli  
131 16.6 66.4 1664976 3 US-09-692-570-1 Sequence 1, Appli  
132 16.6 65.6 1386 3 US-09-328-352-981 Sequence 981, App  
133 16.4 65.6 24000 3 US-09-949-016-14421 Sequence 14421, A  
134 16.4 65.6 92227 3 US-09-949-016-11929 Sequence 11929, A  
135 16.4 65.6 92232 3 US-09-949-016-15421 Sequence 15421, A  
136 16.4 65.6 256171 3 US-09-949-016-12822 Sequence 12822, A  
137 16.4 65.6 256171 3 US-09-949-016-15524 Sequence 15524, A  
138 16.2 64.8 243 3 US-09-513-999C-17532 Sequence 17532, A  
139 16.2 64.8 336 3 US-08-905-223-59 Sequence 59, Appl  
140 16.2 64.8 342 3 US-09-107-532A-2450 Sequence 2450, App  
141 16.2 64.8 601 3 US-09-949-016-71499 Sequence 71499, A  
142 16.2 64.8 601 3 US-09-949-016-71500 Sequence 71500, A  
143 16.2 64.8 601 3 US-09-949-016-91914 Sequence 91914, A  
144 16.2 64.8 601 3 US-09-949-016-165425 Sequence 165425, A  
145 16.2 64.8 601 3 US-09-949-016-165482 Sequence 165482, A  
146 16.2 64.8 601 3 US-09-949-016-194855 Sequence 194855, A  
147 16.2 64.8 601 3 US-09-949-001-685 Sequence 685, App  
148 16.2 64.8 601 3 US-09-949-001-686 Sequence 686, App  
149 16.2 64.8 601 3 US-09-949-002-6221 Sequence 6221, App  
150 16.2 64.8 601 3 US-09-949-002-10285 Sequence 10285, A

ALIGNMENTS

RESULT 1  
US-09-016-434-1010  
; Sequence 1010, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA

; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1010:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 270 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BRAINOT04  
; CLONE: 926642  
US-09-016-434-1010

Query Match 87.2%; Score 21.8; DB 3; Length 270;  
Best Local Similarity 92.0%; Pred. No. 0.46;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAACTGCATCCCAATGTTATG 25  
Db 164 AAAAACTGCATCCCAATGTTATG 188

RESULT 2  
US-09-621-976-14487  
; Sequence 14487, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 14487  
; LENGTH: 460  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-14487

Query Match 87.2%; Score 21.8; DB 3; Length 460;  
Best Local Similarity 92.0%; Pred. No. 0.51;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAACTGCATCCCAATGTTATG 25  
Db 136 AAAAACTGCATCCCAATGTTATG 160

RESULT 3  
US-09-710-279-1467

; Sequence 1467, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1467

; LENGTH: 1917

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-09-710-279-1467

Query Match 81.6%; Score 20.4; DB 3; Length 1917;  
Best Local Similarity 95.5%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAATGCATCCCAATGTTAT 24

Db 1441 AAAAATGCATCCCAATGTTAT 1462

#### RESULT 4

US-09-134-001C-1607

; Sequence 1607, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1607

; LENGTH: 1953

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1607

Query Match 81.6%; Score 20.4; DB 3; Length 1953;  
Best Local Similarity 95.5%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAATGCATCCCAATGTTAT 24

Db 1477 AAAAATGCATCCCAATGTTAT 1498

#### RESULT 5

US-09-710-279-4240

; Sequence 4240, Application US/09710279

; Patent No. 6703492

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/09/710,279

; CURRENT FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4240  
; LENGTH: 3276  
; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-09-710-279-4240

Query Match 81.6%; Score 20.4; DB 3; Length 3276;  
Best Local Similarity 95.5%; Pred. No. 3.9;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAATGCATCCCAATGTTAT 24

Db 1193 AAAAATGCATCCCAATGTTAT 1214

#### RESULT 6

US-09-949-016-179347

; Sequence 179347, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: Fast-SEQ for Windows Version 4.0

; SEQ ID NO 179347

; LENGTH: 601

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-179347

Query Match 76.8%; Score 19.2; DB 3; Length 601;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATGCATCCCAATGTTATG 25

Db 190 AAAAATGCATCCCAATGTTATG 213

#### RESULT 7

US-09-949-016-12386

; Sequence 12386, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: Fast-SEQ for Windows Version 4.0

```
; SEQ ID NO 12386
; LENGTH: 263693
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12386

Query Match          76.8%; Score 19.2; DB 3; Length 263693;
Best Local Similarity 87.5%; Pred. No. 42;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
   |||||
Db 115551 AAAAACTGCATATCCAGTGTATG 115574

RESULT 8
US-09-949-016-16915
; Sequence 63, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 168174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-09-949-016-16915

Query Match          76.8%; Score 19.2; DB 3; Length 263694;
Best Local Similarity 87.5%; Pred. No. 42;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
   |||||
Db 115551 AAAAACTGCATATCCAGTGTATG 115574

RESULT 9
US-10-071-411A-63
; Sequence 63, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Glenn Barnes
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 60/267,515
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 60/314,248
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 168174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(168174)
; OTHER INFORMATION: n = A,T,C or G
US-10-071-411A-63

Query Match          74.4%; Score 18.6; DB 3; Length 168174;
Best Local Similarity 84.0%; Pred. No. 74;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   |||||
Db 91147 AAAGAACTGCATCTTAATGTTCTG 91171

RESULT 10
US-10-071-411A-2
; Sequence 2, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Glenn Barnes
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/10/071,411A
; PRIOR FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 60/267,515
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 60/314,248
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 168273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(168273)
; OTHER INFORMATION: n = A,T,C or G
US-10-071-411A-2

Query Match          74.4%; Score 18.6; DB 3; Length 168273;
Best Local Similarity 84.0%; Pred. No. 74;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   |||||
Db 91246 AAAGAACTGCATCTTAATGTTCTG 91270

RESULT 11
US-09-949-016-11868
; Sequence 11868, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11868
; LENGTH: 300598
; TYPE: DNA
; ORGANISM: Human
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(300598)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11868

Query Match          71.2%; Score 17.8; DB 3; Length 300598;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACTGCATCCCCCAATGTTATG 25
   ||||||| |||||||
Db 205760 AATTGCATCCACAATGTTATG 205780

RESULT 12
US-09-949-016-14588
; Sequence 14588, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14588
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14588

Query Match          71.2%; Score 17.8; DB 3; Length 302604;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACTGCATCCCCCAATGTTATG 25
   ||||||| |||||||
Db 255766 AATTGCATCCACAATGTTATG 255786

RESULT 13
US-09-949-016-14589
; Sequence 14589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14589
```

```
; LENGTH: 302604
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(302604)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14589

Query Match          71.2%; Score 17.8; DB 3; Length 302604;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACTGCATCCCCCAATGTTATG 25
   ||||||| |||||||
Db 255766 AATTGCATCCACAATGTTATG 255786

RESULT 14
US-09-949-016-17119
; Sequence 17119, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17119
; LENGTH: 308362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(308362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17119

Query Match          71.2%; Score 17.8; DB 3; Length 308362;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACTGCATCCCCCAATGTTATG 25
   ||||||| |||||||
Db 255582 AATTGCATCCACAATGTTATG 255602

RESULT 15
US-09-949-016-62866
; Sequence 62866, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14589
```



```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62866
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-62866

Query Match          70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
    ||||| ||||| ||||| ||||| |||||
Db 199 AAAAAATGCATACCAATGGTAT 222

RESULT 16
US-09-489-039A-109/c
; Sequence 109, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 109
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-109

Query Match          70.4%; Score 17.6; DB 3; Length 1770;
Best Local Similarity 83.3%; Pred. No. 82;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
    ||||| ||||| ||||| ||||| |||||
Db 1338 AAAAACTTCATCCCCAGGATATG 1315

RESULT 17
US-09-489-039A-209
; Sequence 209, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 209
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-209

Query Match          70.4%; Score 17.6; DB 3; Length 1803;
Best Local Similarity 83.3%; Pred. No. 82;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
    ||||| ||||| ||||| ||||| |||||
Db 406 AAAAACTTCATCCCCAGGATATG 429

RESULT 18
US-09-949-016-4982
; Sequence 4982, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4982
; LENGTH: 12563
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4982

Query Match          70.4%; Score 17.6; DB 3; Length 12563;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
    ||||| ||||| ||||| ||||| |||||
Db 3191 AGAAAAGACCATCCCAATGTTAT 3214

RESULT 19
US-09-949-016-13553/c
; Sequence 13553, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13553
; LENGTH: 103792
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)..(103792)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13553

Query Match          70.4%; Score 17.6; DB 3; Length 103792;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
    ||||| ||||| ||||| ||||| |||||
Db 54553 AAAAAAATGCATACCCCAATGGTAT 54530
```

RESULT 20  
US-09-949-016-12317/c  
; Sequence 12317, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12317  
; LENGTH: 103993  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(103993)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12317

Query Match 70.4%; Score 17.6; DB 3; Length 103993;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
||||| ||||| ||||| ||||| |||||  
DB 54553 AAAAAATGCATACCAATGGTAT 54530

RESULT 21  
US-09-949-016-16724  
; Sequence 16724, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16724  
; LENGTH: 138693  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16724

Query Match 70.4%; Score 17.6; DB 3; Length 138693;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
||||| ||||| ||||| ||||| |||||  
DB 50685 AGAAAGACCATCCCAATGTTAT 50708

RESULT 22  
US-09-949-016-12287/c  
; Sequence 12287, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12287  
; LENGTH: 285986  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(285986)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12287

Query Match 70.4%; Score 17.6; DB 3; Length 285986;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
||||| ||||| ||||| ||||| |||||  
DB 103371 AAAAACTCAATCTCCAATGTTAT 103348

RESULT 23  
US-09-949-016-14864/c  
; Sequence 14864, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14864  
; LENGTH: 288031  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(288031)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14864

Query Match 70.4%; Score 17.6; DB 3; Length 288031;  
Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
||||| ||||| ||||| ||||| |||||

```
Db 153416 AAGAACTCAATCTCCAATGTTAT 153393

RESULT 24
US-09-620-312D-78/c
; Sequence 78, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-Hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aigong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 78
; LENGTH: 2179
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (284)...(1957)
US-09-620-312D-78

Query Match 69.6%; Score 17.4; DB 3; Length 2179;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 ACTGCATCCCAATGTTAT 24
Db 1272 ACTGCAACCCCAATGTTAT 1254

RESULT 25
US-09-270-767-11666
; Sequence 11666, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11666
; LENGTH: 881
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11666

Query Match 68.8%; Score 17.2; DB 3; Length 881;

Db 153416 AAGAACTCAATCTCCAATGTTAT 153393

Best Local Similarity 86.4%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
Db 738 AAAACTGAATCAACAATGTTAT 759

RESULT 26
US-09-345-236B-83
; Sequence 83, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tuku, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 3198
; TYPE: DNA
; ORGANISM: mosquito baculovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2545)...(3198)
; NAME/KEY: CDS
; LOCATION: (742)...(1056)
; NAME/KEY: CDS
; LOCATION: (1837)...(1998)
US-09-345-236B-83

Query Match 68.8%; Score 17.2; DB 3; Length 3198;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
Db 1279 AAAATTGCATCCCAATGTTAT 1300

RESULT 27
US-09-345-236B-87/c
; Sequence 87, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tuku, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 3198
; TYPE: DNA
; ORGANISM: mosquito baculovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (428)...(3195)
```

```
; NAME/KEY: CDS
; LOCATION: (20)....(241)
US-09-345-236B-87

Query Match      68.8%; Score 17.2; DB 3; Length 3198;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AAAAAGTCATCCCAATGTTAT 24
      ||||| ||||| ||||| |||||
Db      1920 AAAAATGCATCCCAATGTTAT 1899

RESULT 28
US-09-345-236B-90/c
; Sequence 90, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tokuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 90
; LENGTH: 3198
; TYPE: DNA
; ORGANISM: mosquito baculovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (667)....(897)
US-09-345-236B-90

Query Match      68.8%; Score 17.2; DB 3; Length 3198;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AAAAAGTCATCCCAATGTTAT 24
      ||||| ||||| ||||| |||||
Db      1920 AAAAATGCATCCCAATGTTAT 1899

RESULT 29
US-09-345-236B-92/c
; Sequence 92, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tokuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 92
; LENGTH: 3198
; TYPE: DNA
; ORGANISM: mosquito baculovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (667)....(897)
US-09-345-236B-92

; LOCATION: (696)....(848)
US-09-345-236B-92

Query Match      68.8%; Score 17.2; DB 3; Length 3198;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AAAAAGTCATCCCAATGTTAT 24
      ||||| ||||| ||||| |||||
Db      1920 AAAAATGCATCCCAATGTTAT 1899

RESULT 30
US-08-836-329-1
; Sequence 1, Application US/08836329
; Patent No. 6090546
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Method for the Detection of Ras Oncogenes,
; TITLE OF INVENTION: In Particular The K-Ras Oncogene
; NUMBER OF SEQUENCES: 16
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,329
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3918 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 825..947
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1388..1567
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2118..2278
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3034..3158
; NAME/KEY: exon
; LOCATION: 3459..3616
; US-08-836-329-1

Query Match      68.8%; Score 17.2; DB 3; Length 3918;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AAAAAGTCATCCCAATGTTAT 24
      ||||| ||||| ||||| |||||
Db      2879 AAAAATGCATCCCAATGTTAT 2900

RESULT 31
US-09-949-016-17001
; Sequence 17001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17001  
; LENGTH: 22605  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(22605)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17001

Query Match 68.8%; Score 17.2; DB 3; Length 22605;  
Best Local Similarity 86.4%; Pred. No. 2.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGTGCATCCCAATGTTAT 24  
|||||  
Db 13612 AAAAAGTGCATCCCAATGTTAT 13633

RESULT 32  
US-09-949-016-12032/c  
; Sequence 12032, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12032  
; LENGTH: 40655  
; TYPE: DNA  
; ORGANISM: Human  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12032

Query Match 68.8%; Score 17.2; DB 3; Length 40655;  
Best Local Similarity 86.4%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGTGCATCCCAATGTT 22  
|||||  
Db 29026 AAAAAGTGCATCCCAATGTT 29005

RESULT 33  
US-09-949-016-15919/c  
; Sequence 15919, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15919  
; LENGTH: 40655  
; TYPE: DNA  
; ORGANISM: Human  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15919

Query Match 68.8%; Score 17.2; DB 3; Length 40655;  
Best Local Similarity 86.4%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGTGCATCCCAATGTT 22  
|||||  
Db 29026 AAAAAGTGCATCCCAATGTT 29005

RESULT 34  
US-09-949-016-12444  
; Sequence 12444, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12444  
; LENGTH: 41696  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(41696)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12444

Query Match 68.8%; Score 17.2; DB 3; Length 41696;  
Best Local Similarity 86.4%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGTGCATCCCAATGTTAT 24  
|||||  
Db 13612 AAAAAGTGCATCCCAATGTTAT 13633

RESULT 35  
US-09-949-016-16851  
; Sequence 16851, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16851  
; LENGTH: 58782  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16851

Query Match 68.8%; Score 17.2; DB 3; Length 58782;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTCATCCCAATGTTA 23  
||||| |||||||  
DB 47134 AAAAAGTCATCCCAATGTTA 47155

RESULT 36  
US-09-949-016-16553/c  
; Sequence 16553, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16553  
; LENGTH: 94142  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)..(94142)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16553

Query Match 68.8%; Score 17.2; DB 3; Length 94142;  
Best Local Similarity 86.4%; Pred. No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAAGTCATCCCAATGTTATG 25  
||||| |||||||  
DB 55395 AAAGTCATCCCAATGTTATG 55374

RESULT 37  
US-09-949-016-16723  
; Sequence 16723, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16723  
; LENGTH: 157822  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16723

Query Match 68.8%; Score 17.2; DB 3; Length 157822;  
Best Local Similarity 86.4%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAGTCATCCCAATGTTAT 24  
||||| |||||||  
DB 86913 AAAAGTCATCCCAATGTTAT 86934

RESULT 38  
US-09-513-999C-16990/c  
; Sequence 16990, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 16990  
; LENGTH: 134  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; LOCATION: 18  
; OTHER INFORMATION: w=a or t  
US-09-513-999C-16990

Query Match 68.0%; Score 17; DB 3; Length 134;  
Best Local Similarity 80.0%; Pred. No. 90;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAGTCATCCCAATGTTATG 25  
||||| |||||||  
DB 68 ATAAACTATACCCCAATTTTIG 44

RESULT 39  
US-09-513-999C-22791/c  
; Sequence 22791, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26

```
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 22791
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6
; OTHER INFORMATION: h=a or c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 160
; OTHER INFORMATION: w=a or t
; OTHER INFORMATION: 22791
US-09-513-999C-22791

Query Match      68.0%; Score 17; DB 3; Length 171;
Best Local Similarity 80.0%; Pred. No. 95;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 41 AAAAAAATCTATCGCAATGTTATG 17

RESULT 40
US-09-270-767-6867/c
; Sequence 6867, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6867
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6867

Query Match      68.0%; Score 17; DB 3; Length 173;
Best Local Similarity 80.0%; Pred. No. 95;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 112 AAACAACTGCAACCGCAACGTTGTG 88

RESULT 41
US-09-270-767-22149/c
; Sequence 22149, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6867
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-22149/c

Query Match      68.0%; Score 17; DB 3; Length 173;
Best Local Similarity 80.0%; Pred. No. 95;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 112 AAACAACTGCAACCGCAACGTTGTG 88

RESULT 42
US-09-134-000C-150/c
; Sequence 150, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-150

Query Match      68.0%; Score 17; DB 3; Length 243;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 67 AAATATCTGCATCCCAATGATAGG 43

RESULT 43
US-09-270-767-4566
; Sequence 4566, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4566
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-4566

Query Match      68.0%; Score 17; DB 3; Length 430;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 312 AAAGAACTGCATCGCAATATCATG 336
```







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-25

Perfect score: 25

Sequence: 1 aaaaactgcaccccaatgttatg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	US-10-719-900-25
2	24	96.0	25	7	US-10-719-956-283109
3	23.4	93.6	25	8	US-10-719-900-26
4	22.4	89.6	25	7	US-10-719-956-283110
5	22.4	89.6	3977	6	US-10-400-435A-1
6	21.8	87.2	270	6	US-10-305-720-1010
7	20.4	81.6	1914	7	US-10-282-122A-34587
8	20.4	81.6	1953	7	US-10-724-972A-1131
9	20.2	80.8	619	7	US-10-424-599-111654
10	19.8	79.2	1904	7	US-10-424-599-111652
11	19.2	76.8	619	4	US-09-925-065A-510687
12	18.8	75.2	675	4	US-09-925-065A-905792
13	18.8	75.2	53542	3	US-09-801-574-61
14	18.6	74.4	168174	7	US-10-071-411-63
15	18.6	74.4	168174	8	US-10-914-799-63
16	18.6	74.4	168273	7	US-10-071-411-2
17	18.6	74.4	168273	8	US-10-914-799-2
18	18.2	72.8	288	7	US-10-424-599-2458
19	18.2	72.8	488	4	US-09-925-065A-603866
20	18.2	72.8	488	4	US-09-925-065A-603867
21	18.2	72.8	488	4	US-09-925-065A-603868
22	18.2	72.8	545	4	US-09-925-065A-133035
23	18.2	72.8	580	4	US-09-925-065A-788458

Sequence 367029, A	588	72.8	18.2	C	24	US-09-925-065A-367029
Sequence 38192, A	618	72.8	18.2	C	25	US-09-925-065A-38192
Sequence 427422, A	633	72.8	18.2	C	26	US-09-925-065A-427422
Sequence 6757, App	13080	72.8	18.2	C	27	US-10-719-993-6757
Sequence 160, App	95683	72.8	18.2	C	28	US-10-087-192-160
Sequence 5682, App	561515	72.8	18.2	C	29	US-10-741-601-5682
Sequence 17730, A	561515	72.8	18.2	C	30	US-10-741-600-17730
Sequence 2455, App	467	71.2	17.8	C	31	US-10-424-599-2455
Sequence 133034, A	545	71.2	17.8	C	32	US-09-925-065A-133034
Sequence 271268, A	568	71.2	17.8	C	33	US-09-925-065A-271268
Sequence 271269, A	568	71.2	17.8	C	34	US-09-925-065A-271269
Sequence 271270, A	568	71.2	17.8	C	35	US-09-925-065A-271270
Sequence 271271, A	568	71.2	17.8	C	36	US-09-925-065A-271271
Sequence 18209, A	578	71.2	17.8	C	37	US-10-363-345A-18209
Sequence 18210, A	578	71.2	17.8	C	38	US-10-363-345A-18210
Sequence 18209, A	578	71.2	17.8	C	39	US-10-363-483A-18209
Sequence 18210, A	578	71.2	17.8	C	40	US-10-363-483A-18210
Sequence 427423, A	633	71.2	17.8	C	41	US-09-925-065A-427423
Sequence 12160, A	1083	71.2	17.8	C	42	US-10-282-122A-12160
Sequence 668046, A	3517	71.2	17.8	C	43	US-09-925-065A-668046
Sequence 89164, A	4283	71.2	17.8	C	44	US-10-437-963-89164
Sequence 6990, App	303172	71.2	17.8	C	45	US-10-719-993-6990
Sequence 6234, App	379	70.4	17.6	C	46	US-10-424-599-6234
Sequence 62193, A	389	70.4	17.6	C	47	US-10-425-115-62193
Sequence 26833, A	461	70.4	17.6	C	48	US-09-918-995-26833
Sequence 464930, A	519	70.4	17.6	C	49	US-09-925-065A-464930
Sequence 180366, A	520	70.4	17.6	C	50	US-10-027-632-180366
Sequence 180366, A	520	70.4	17.6	C	51	US-10-027-632-180366
Sequence 395978, A	527	70.4	17.6	C	52	US-09-925-065A-395978
Sequence 395979, A	527	70.4	17.6	C	53	US-09-925-065A-395979
Sequence 880165, A	570	70.4	17.6	C	54	US-09-925-065A-880165
Sequence 313962, A	582	70.4	17.6	C	55	US-09-925-065A-313962
Sequence 895797, A	593	70.4	17.6	C	56	US-09-925-065A-895797
Sequence 1132, App	596	70.4	17.6	C	57	US-09-764-860-1132
Sequence 1132, App	596	70.4	17.6	C	58	US-10-074-095-1132
Sequence 56255, A	596	70.4	17.6	C	59	US-10-312-872-1132
Sequence 364719, A	612	70.4	17.6	C	60	US-10-927-930-56255
Sequence 364720, A	613	70.4	17.6	C	61	US-09-925-065A-364719
Sequence 268, App	614	70.4	17.6	C	62	US-09-925-065A-364720
Sequence 268, App	614	70.4	17.6	C	63	US-09-764-860-268
Sequence 288, App	614	70.4	17.6	C	64	US-10-074-095-268
Sequence 798669, A	614	70.4	17.6	C	65	US-10-212-872-268
Sequence 230053, A	618	70.4	17.6	C	66	US-09-925-065A-798669
Sequence 230053, A	624	70.4	17.6	C	67	US-10-027-632-230053
Sequence 230053, A	624	70.4	17.6	C	68	US-10-027-632-230053
Sequence 54943, A	624	70.4	17.6	C	69	US-10-425-115-54943
Sequence 41816, A	624	70.4	17.6	C	70	US-10-282-122A-41816
Sequence 55232, A	1086	70.4	17.6	C	71	US-09-925-065A-55232
Sequence 30931, A	1421	70.4	17.6	C	72	US-10-282-122A-30931
Sequence 839, App	1506	70.4	17.6	C	73	US-10-767-795-839
Sequence 101593, A	2502	70.4	17.6	C	74	US-10-027-632-101593
Sequence 101593, A	2502	70.4	17.6	C	75	US-10-027-632-101593
Sequence 68, Appl	2975	70.4	17.6	C	76	US-10-320-797-68
Sequence 12, Appl	3833	70.4	17.6	C	77	US-10-338-777-12
Sequence 2903, App	4963	70.4	17.6	C	78	US-09-764-877-2903
Sequence 2903, App	4963	70.4	17.6	C	79	US-10-242-515-2903
Sequence 1182, App	10256	70.4	17.6	C	80	US-09-764-847-1182
Sequence 2567, App	10256	70.4	17.6	C	81	US-10-092-154-1182
Sequence 235, App	12563	70.4	17.6	C	82	US-10-723-860-2567
Sequence 6713, App	12563	70.4	17.6	C	83	US-10-745-237-255
Sequence 3, Appl	12775	70.4	17.6	C	84	US-10-723-860-6713
Sequence 17745, A	68200	70.4	17.6	C	85	US-10-840-590-3
Sequence 10, Appl	244196	70.4	17.6	C	86	US-10-741-600-17745
Sequence 2058, App	1163020	70.4	17.6	C	87	US-10-398-221-10
Sequence 1, Appl	3011208	70.4	17.6	C	88	US-10-398-221-2058
Sequence 354661, A	3673778	70.4	17.6	C	89	US-10-312-841-1
Sequence 78, Appl	594	69.6	17.4	C	90	US-09-925-065A-354661
Sequence 78, Appl	2179	69.6	17.4	C	91	US-10-037-270-78
Sequence 78, Appl	2179	69.6	17.4	C	92	US-10-117-722-78
Sequence 19906, A	2179	69.6	17.4	C	93	US-10-122-851-78
Sequence 574, App	3790	69.6	17.4	C	94	US-09-814-353-19906
Sequence 31318, A	17528	69.6	17.4	C	95	US-10-311-455-574
	201	68.8	17.2	C	96	US-10-719-993-31318

```

97 17.2 68.8 311 8 US-10-674-124A-26606
98 17.2 68.8 427 8 US-10-767-795-1414
c 99 17.2 68.8 451 7 US-10-437-963-91182
100 17.2 68.8 495 8 US-10-425-115-53972
c 101 17.2 68.8 513 7 US-10-437-963-52090
102 17.2 68.8 570 4 US-09-925-065A-880166
103 17.2 68.8 620 4 US-09-925-065A-341098
104 17.2 68.8 624 4 US-09-925-065A-341099
c 105 17.2 68.8 624 4 US-09-925-065A-736692
106 17.2 68.8 634 4 US-09-925-065A-745337
c 107 17.2 68.8 634 5 US-10-027-632-240922
108 17.2 68.8 645 5 US-10-027-632-240923
109 17.2 68.8 645 5 US-10-027-632-240924
110 17.2 68.8 645 5 US-10-027-632-240925
111 17.2 68.8 645 5 US-10-027-632-240922
112 17.2 68.8 645 5 US-10-027-632-240922
113 17.2 68.8 645 6 US-10-027-632-240924
114 17.2 68.8 645 6 US-10-027-632-240924
115 17.2 68.8 645 6 US-10-027-632-240925
c 116 17.2 68.8 681 7 US-10-282-122A-10935
c 117 17.2 68.8 975 5 US-10-027-632-116833
c 118 17.2 68.8 975 6 US-10-027-632-116833
c 119 17.2 68.8 1012 9 US-10-499-184-11
c 120 17.2 68.8 1039 5 US-10-027-632-121527
c 121 17.2 68.8 1039 5 US-10-027-632-121528
c 122 17.2 68.8 1039 5 US-10-027-632-121529
c 123 17.2 68.8 1039 6 US-10-027-632-121527
c 124 17.2 68.8 1039 6 US-10-027-632-121528
c 125 17.2 68.8 1039 6 US-10-027-632-121529
c 126 17.2 68.8 1875 7 US-10-260-238-903
c 127 17.2 68.8 7422 10 US-11-097-143-10606
c 128 17.2 68.8 7447 10 US-11-097-143-25978
c 129 17.2 68.8 8132 10 US-11-097-143-25981
c 130 17.2 68.8 36714 7 US-10-322-281-864
c 131 17.2 68.8 38605 5 US-10-087-192-304
c 132 17.2 68.8 62231 3 US-10-087-192-493
c 133 17.2 68.8 96588 3 US-09-997-722-292
c 134 17.2 68.8 268685 6 US-10-265-071-22
c 135 17.2 68.8 268685 6 US-10-025-966A-22
c 136 17.2 68.8 775062 8 US-10-719-993-6844
c 137 17 68.0 201 8 US-10-719-993-8989
c 138 17 68.0 201 8 US-10-719-993-8991
c 139 17 68.0 201 8 US-10-719-993-8992
c 140 17 68.0 201 8 US-10-719-993-8993
c 141 17 68.0 409 4 US-09-925-065A-172463
c 142 17 68.0 410 4 US-09-925-065A-532812
c 143 17 68.0 426 5 US-10-027-632-182442
c 144 17 68.0 426 5 US-10-027-632-307207
c 145 17 68.0 426 6 US-10-027-632-182442
c 146 17 68.0 426 6 US-10-027-632-307207
c 147 17 68.0 427 5 US-10-027-632-79765
c 148 17 68.0 427 5 US-10-027-632-109526
c 149 17 68.0 427 6 US-10-027-632-79765
c 150 17 68.0 427 6 US-10-027-632-109526

Sequence 26606, A
Sequence 1414, Ap
Sequence 91182, A
Sequence 53972, A
Sequence 52090, A
Sequence 880166,
Sequence 341098,
Sequence 341099,
Sequence 736692,
Sequence 745089,
Sequence 745337,
Sequence 240922,
Sequence 240923,
Sequence 240924,
Sequence 240925,
Sequence 10935, A
Sequence 116833,
Sequence 116833,
Sequence 11, Ap1
Sequence 121527,
Sequence 121528,
Sequence 121529,
Sequence 121527,
Sequence 121528,
Sequence 121529,
Sequence 903, Ap
Sequence 10606, A
Sequence 25978, A
Sequence 25981, A
Sequence 864, Ap
Sequence 304, Ap
Sequence 493, Ap
Sequence 292, Ap
Sequence 22, Ap1
Sequence 22, Ap1
Sequence 6844, Ap
Sequence 8989, Ap
Sequence 8991, Ap
Sequence 8992, Ap
Sequence 8993, Ap
Sequence 172463,
Sequence 592812,
Sequence 182442,
Sequence 307207,
Sequence 182442,
Sequence 307207,
Sequence 79765, A
Sequence 109526,
Sequence 79765, A
Sequence 109526,

; SEQ ID NO 25
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-25

Query Match 100.0%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.17; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
Db 1 AAAAACTGCATCCCAATGTTATG 25

RESULT 2
US-10-719-956-283109
; Sequence 283109, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,836
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 699466
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 283109
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-719-956-283109

Query Match 96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
Db 2 AAAAACTGCATCCCAATGTTATG 25

RESULT 3
US-10-719-900-26
; Sequence 26, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 26
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-26

Query Match 93.6%; Score 23.4; DB 8; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.93;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
Db 1 AAAAACTGCATCCCAATGTTATG 25

ALIGNMENTS

RESULT 1
US-10-719-900-25
; Sequence 25, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002 11 20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

## RESULT 4

US-10-719-956-283110  
; Sequence 283110, Application US/10719956  
; Publication No. US20040146910A1  
; GENERAL INFORMATION:  
; APPLICANT: Xue Mei Zhou  
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat  
; FILE REFERENCE: 3527.1  
; CURRENT APPLICATION NUMBER: US/10/719,956  
; CURRENT FILING DATE: 2003-11-20  
; PRIOR APPLICATION NUMBER: 60/427,836  
; PRIOR FILING DATE: 2002 11 20  
; NUMBER OF SEQ ID NOS: 699466  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 283110  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-10-719-956-283110

Query Match 89.6%; Score 22.4; DB 7; Length 25;  
Best Local Similarity 95.8%; Pred. No. 2.7;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAGTCATCCCAATGTTATG 25  
|||||  
Db 2 AAAAAGTCATCCCAATGTTATG 25

## RESULT 5

US-10-400-435A-1  
; Sequence 1, Application US/10400435A  
; Publication No. US20030224416A1  
; GENERAL INFORMATION:  
; APPLICANT: Cochran, Susan  
; APPLICANT: Yamagami, Keiji  
; APPLICANT: Ohashi, Yoshitaka  
; TITLE OF INVENTION: SCHIZOPHRENIA RELATED GENE  
; FILE REFERENCE: 2003-0437A/WNC/00279  
; CURRENT APPLICATION NUMBER: US/10/400,435A  
; CURRENT FILING DATE: 2003-03-28  
; PRIOR APPLICATION NUMBER: EP02007114.8  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 3977  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1162)..(2919)  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (1)..(3977)  
; FEATURE:  
; NAME/KEY: polyA\_signal  
; LOCATION: (3958)..(3963)  
; PUBLICATION INFORMATION:  
; AUTHORS: Lunau C.J., Williams, J.B., Marshall, J., Levitan, E.S., Oliva, C.,  
; AUTHORS: Smith, J.S., Antanavage, J., Folander, K., Stein, R.B., Swanson, R., Kaczmarek, L.  
; AUTHORS: and Buhrow, S.A.  
; TITLE: Alternate splicing contributes to K+ channel diversity in the mammalian  
; TITLE: central nervous system.  
; JOURNAL: Proc. Natl. Acad. Sci. USA.  
; VOLUME: 88  
; ISSUE: 5  
; PAGES: 3932-3936  
; DATABASE ACCESSION NUMBER: NM\_012856  
; DATABASE ENTRY DATE: 1991-05-01  
US-10-400-435A-1

Query Match 89.6%; Score 22.4; DB 6; Length 3977;

Best Local Similarity 95.8%; Pred. No. 6.7;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAGTCATCCCAATGTTATG 25  
|||||  
Db 3915 AAAAAGTCATCCCAATGTTATG 3938

## RESULT 6

US-10-305-720-1010  
; Sequence 1010, Application US/10305720  
; Publication No. US20040010136A1  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.  
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression  
; FILE REFERENCE: PA-0002-1 CON  
; CURRENT APPLICATION NUMBER: US/10/305,720  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/016,434  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 1490  
; SOFTWARE: PERL Program  
; SEQ ID NO 1010  
; LENGTH: 270  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20040010136A1 926642  
US-10-305-720-1010

Query Match 87.2%; Score 21.8; DB 6; Length 270;  
Best Local Similarity 92.0%; Pred. No. 7.8;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAGTCATCCCAATGTTATG 25  
|||||  
Db 164 AAAAAGTCATCCCAATGTTATG 188

## RESULT 7

US-10-282-122A-34587  
; Sequence 34587, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34587
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-34587

Query Match      81.6%; Score 20.4; DB 7; Length 1914;
Best Local Similarity 95.5%; Pred. No. 49;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
    |||||
Db 1441 AAAACTGCATCCTCAATGTTAT 1462

RESULT 8
US-10-724-972A-1131
; Sequence 1131, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 1131
; LENGTH: 1953
; TYPE: DNA
; ORGANISM: S.epidermidis
US-10-724-972A-1131

Query Match      81.6%; Score 20.4; DB 7; Length 1953;
Best Local Similarity 95.5%; Pred. No. 49;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAACTGCATCCCAATGTTAT 24
    |||||
Db 1477 AAAACTGCATCCTCAATGTTAT 1498

RESULT 9
US-10-424-599-111654
; Sequence 111654, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 111652
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71831C.1
US-10-424-599-111652

Query Match      79.2%; Score 19.8; DB 7; Length 1904;
Best Local Similarity 91.3%; Pred. No. 92;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCATCCCAATGTTA 23
    |||||
Db 1292 ATAAAAATGCATCCCAATGTTA 1314

RESULT 11
US-09-925-065A-510687
; Sequence 510687, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

```
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 111654
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71833C.1
US-10-424-599-111654

Query Match      80.8%; Score 20.2; DB 7; Length 619;
Best Local Similarity 88.0%; Pred. No. 49;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGTCATCCCAATGTTATG 25
    |||||
Db 45 ATAAAAATGCATCCCAATGTTAAG 69

RESULT 10
US-10-424-599-111652
; Sequence 111652, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 111652
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71831C.1
US-10-424-599-111652

Query Match      79.2%; Score 19.8; DB 7; Length 1904;
Best Local Similarity 91.3%; Pred. No. 92;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCATCCCAATGTTA 23
    |||||
Db 1292 ATAAAAATGCATCCCAATGTTA 1314

RESULT 11
US-09-925-065A-510687
; Sequence 510687, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 510687  
; LENGTH: 619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-510687

Query Match 76.8%; Score 19.2; DB 4; Length 619;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24  
Db 142 AAAAAATATGATCCCAATGTTAT 165

RESULT 12  
US-09-925-065A-905792  
; Sequence 905792, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 905792  
; LENGTH: 675  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-905792

Query Match 75.2%; Score 18.8; DB 4; Length 675;  
Best Local Similarity 90.9%; Pred. No. 2.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTT 22  
Db 530 AATAAACTGCATCCCAATTT 551

RESULT 13  
US-09-801-574-61  
; Sequence 61, Application US/09801574  
; Patent No. US20020081592A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Peijing Jeremy  
; APPLICANT: Page, David C.  
; TITLE OF INVENTION: Reproduction-Specific Genes  
; FILE REFERENCE: 0399.2007-002  
; CURRENT APPLICATION NUMBER: US/09/801,574  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: 60/187,518  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: 60/261,557  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 61  
; LENGTH: 53542  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-801-574-61

Query Match 75.2%; Score 18.8; DB 3; Length 53542;  
Best Local Similarity 90.9%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAATGTCATCCCAATGTTATG 25  
Db 29516 AAATTCATCCCAATGATG 29537

RESULT 14  
US-10-071-411-63  
; Sequence 63, Application US/10071411  
; Publication No. US20040170974A1  
; GENERAL INFORMATION:  
; APPLICANT: Joanne Meyer  
; TITLE OF INVENTION: Detection of Polymorphisms in the Human  
; FILE REFERENCE: MRI-021  
; CURRENT APPLICATION NUMBER: US/10/071,411  
; CURRENT FILING DATE: 2002-02-07  
; PRIOR APPLICATION NUMBER: 60/267,515  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 60/314,248  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63  
; LENGTH: 168174  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(168174)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-071-411-63

Query Match 74.4%; Score 18.6; DB 7; Length 168174;  
Best Local Similarity 84.0%; Pred. No. 7.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25  
Db 91147 AAAGAACTGCATCCTTAATGTTCTG 91171

RESULT 15  
US-10-914-799-63  
; Sequence 63, Application US/10914799  
; Publication No. US20050009084A1  
; GENERAL INFORMATION:  
; APPLICANT: Joanne Meyer  
; TITLE OF INVENTION: Detection of Polymorphisms in the Human  
; FILE REFERENCE: MRI-021  
; CURRENT APPLICATION NUMBER: US/10/914,799  
; CURRENT FILING DATE: 2004-08-09  
; PRIOR APPLICATION NUMBER: US/10/071,411  
; PRIOR FILING DATE: 2002-02-07  
; PRIOR APPLICATION NUMBER: 60/267,515  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 60/314,248  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63





```
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 603866
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-603866
```

```
Query Match          72.8%; Score 18.2; DB 4; Length 488;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAACTGCATCCCAATGTTA 23
      ||||| ||||| ||||| |||||
Db 40 AAAAAAAGCATGCCCAATGTTA 62
```

## RESULT 20

```
US-09-925-065A-603867
; Sequence 603867, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 603867
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-603867
```

```
Query Match          72.8%; Score 18.2; DB 4; Length 488;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAACTGCATCCCAATGTTA 23
      ||||| ||||| ||||| |||||
Db 40 AAAAAAAGCATGCCCAATGTTA 62
```

## RESULT 21

```
US-09-925-065A-603868
; Sequence 603868, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
```

```
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 603868
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-603868
```

```
Query Match          72.8%; Score 18.2; DB 4; Length 488;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAACTGCATCCCAATGTTA 23
      ||||| ||||| ||||| |||||
Db 40 AAAAAAAGCATGCCCAATGTTA 62
```

## RESULT 22

```
US-09-925-065A-133035/c
; Sequence 133035, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133035
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-133035
```

```
Query Match          72.8%; Score 18.2; DB 4; Length 545;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 AAAAACTGCATCCCAATGTTAT 24
      ||||| ||||| ||||| |||||
Db 197 AAAAACTGCATCCCAAAAGTAAT 175
```

## RESULT 23

```
US-09-925-065A-788458
; Sequence 788458, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
```

```
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 60/261,766
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 788458
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-788458

Query Match          72.8%; Score 18.2; DB 4; Length 580;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCATGCCCAATGTTA 23
Db 321 AAAAAAAGCATGCCCAATGTTA 343

RESULT 24
US-09-925-065A-367029/c
; Sequence 367029, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 367029
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-367029

Query Match          72.8%; Score 18.2; DB 4; Length 588;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCATGCCCAATGTTA 23
Db 301 AAAAAAAGCATGCCCAATGTTA 279

RESULT 25
US-09-925-065A-38192
; Sequence 38192, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427422
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-427422/c
; Sequence 427422, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427422
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-427422

Query Match          72.8%; Score 18.2; DB 4; Length 633;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCATGCCCAATGTTA 23
Db 299 AAAAAAAGCATGCCCAATGTTA 277

RESULT 27
US-10-719-993-6757/c
; Sequence 6757, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38192
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-38192

Query Match          72.8%; Score 18.2; DB 4; Length 618;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCATGCCCAATGTTA 23
Db 353 AAAAAAAGCATGCCCAATGTTA 375

RESULT 26
US-09-925-065A-427422/c
; Sequence 427422, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427422
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-427422

Query Match          72.8%; Score 18.2; DB 4; Length 633;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGCATGCCCAATGTTA 23
Db 299 AAAAAAAGCATGCCCAATGTTA 277

RESULT 27
US-10-719-993-6757/c
; Sequence 6757, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
```

FILE REFERENCE: CL001496  
CURRENT APPLICATION NUMBER: US/10/719,993  
CURRENT FILING DATE: 2003-11-24  
NUMBER OF SEQ ID NOS: 5342  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6757  
LENGTH: 13080  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-719-993-6757

Query Match 72.8%; Score 18.2; DB 8; Length 13080;  
Best Local Similarity 87.0%; Pred. No. 7.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTA 23  
|||||  
DB 3685 AAAACACTGCTTCCCAATGTTA 3663  
|||||

## RESULT 28

US-10-087-192-160  
Sequence 160, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:

APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: CANCER  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 160  
LENGTH: 95683  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-087-192-160

Query Match 72.8%; Score 18.2; DB 5; Length 95683;  
Best Local Similarity 87.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTAT 24  
|||||  
DB 8156 AAATACTGCAACCCCTATGTTAT 8178  
|||||

## RESULT 29

US-10-741-601-5682/c  
Sequence 5682, Application US/10741601  
Publication No. US20040168519A1  
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001500  
CURRENT APPLICATION NUMBER: US/10/741,601  
CURRENT FILING DATE: 2003-12-22  
NUMBER OF SEQ ID NOS: 26415  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5682  
LENGTH: 561515  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(561515)

OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-741-601-5682

Query Match 72.8%; Score 18.2; DB 7; Length 561515;  
Best Local Similarity 87.0%; Pred. No. 1.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTA 23  
|||||  
DB 253561 AAAAACCTCTCTCCCAATGTTA 253539  
|||||

## RESULT 30

US-10-741-600-17730/c  
Sequence 17730, Application US/10741600  
Publication No. US20050026169A1  
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001499  
CURRENT APPLICATION NUMBER: US/10/741,600  
CURRENT FILING DATE: 2003-12-22  
NUMBER OF SEQ ID NOS: 73997  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17730  
LENGTH: 561515  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(561515)  
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-10-741-600-17730

Query Match 72.8%; Score 18.2; DB 8; Length 561515;  
Best Local Similarity 87.0%; Pred. No. 1.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTA 23  
|||||  
DB 253561 AAAAACCTCTCTCCCAATGTTA 253539  
|||||

## RESULT 31

US-10-424-599-2455  
Sequence 2455, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:

APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 2455  
LENGTH: 467  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_102222C.1  
US-10-424-599-2455

Query Match 71.2%; Score 17.8; DB 7; Length 467;

Best Local Similarity 90.5%; Pred. No. 6e+02; Indels 0; Gaps 0;

QY 4 AAAACTGCATCCCAATGTTAT 24  
|||||

```
Db      399 AAAGTGCATCCCAATTTT 419

RESULT 32
US-09-925-065A-133034/c
; Sequence 133034, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133034
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-133034

Query Match      71.2%; Score 17.8; DB 4; Length 545;
Best Local Similarity 82.6%; Pred. No. 6.1e+02;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAGTGCATCCCAATGTT 24
Db      197 AAAGTGCATCCCAAGTAA 175

RESULT 33
US-09-925-065A-271268/c
; Sequence 271268, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271268
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-271268

Query Match      71.2%; Score 17.8; DB 4; Length 568;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAGTGCATCCCAATGTT 22
Db      259 AAAGTGCATCCCAATGTT 239

RESULT 34
US-09-925-065A-271269/c
; Sequence 271269, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271269
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-271269

Query Match      71.2%; Score 17.8; DB 4; Length 568;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAGTGCATCCCAATGTT 22
Db      259 AAAGTGCATCCCAATGTT 239

RESULT 35
US-09-925-065A-271270/c
; Sequence 271270, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271270
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-271270
```

```
Query Match          71.2%; Score 17.8; DB 4; Length 568;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGT 22
   ||||| ||||| |||||
Db 259 AAAAACTCTCCCAATGT 239

RESULT 36
US-09-925-065A-271271/c
; Sequence 271271, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271271
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-271271

Query Match          71.2%; Score 17.8; DB 4; Length 568;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGCATCCCAATGT 22
   ||||| ||||| |||||
Db 259 AAAAACTCTCCCAATGT 239

RESULT 37
US-10-363-345A-18209/c
; Sequence 18209, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18209
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18209
US-10-363-345A-18209

Query Match          71.2%; Score 17.8; DB 8; Length 578;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAACTGCATCCCAATGT 21
   ||||| ||||| |||||
Db 403 AAAAAACCGCATCCCAATTT 383

RESULT 38
US-10-363-345A-18210
; Sequence 18210, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18210
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18210
US-10-363-345A-18210

Query Match          71.2%; Score 17.8; DB 8; Length 578;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACTGCATCCCAATGT 21
   ||||| ||||| |||||
Db 176 AAAAAACCGCATCCCAATTT 196

RESULT 39
US-10-363-483A-18209/c
; Sequence 18209, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18209
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18209
US-10-363-483A-18209

Query Match          71.2%; Score 17.8; DB 9; Length 578;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAACTGCATCCCAATGT 21
   ||||| ||||| |||||
Db 403 AAAAAACCGCATCCCAATTT 383

RESULT 40
US-10-363-483A-18210
; Sequence 18210, Application US/10363483A
```

```
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 18210
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 18210
US-10-363-483A-18210

Query Match          71.2%; Score 17.8; DB 9; Length 578;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAACTGCATCCCAATGT 21
        ||||| ||||| ||||| |||||
Db       176 AAAAAACCGCATCCCAATTT 196

RESULT 41
US-09-925-065A-427423/C
; Sequence 427423, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427423
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-427423

Query Match          71.2%; Score 17.8; DB 4; Length 633;
Best Local Similarity 82.6%; Pred. No. 6.3e+02;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAACTGCATCCCAATGTTA 23
        ||||| ||||| ||||| |||||
Db       299 AAAAAAAGCATGCCCAATGTTA 277

RESULT 42
US-10-282-122A-12160/c
; Sequence 12160, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12160
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Bacteroides fragilis
US-10-282-122A-12160

Query Match          71.2%; Score 17.8; DB 7; Length 1083;
Best Local Similarity 90.5%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAAACTGCATCCCAATGTT 22
        ||||| ||||| ||||| |||||
Db       633 AAAAACTGCATCCCAATCTT 613

RESULT 43
US-09-925-065A-668046/c
; Sequence 668046, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```



```

; PRIOR FILING DATE: 2001-05-09
;
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 668046
; LENGTH: 3517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-668046

```

```
Query Match          71.2%; Score 17.8; DB 4; Length 3517;
Best Local Similarity 90.5%; Pred. No. 8.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

RESULT 44
US-10-437-963-89164
; Sequence 89164, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION: Thomas J.
; APPLICANT: La Rosa, David K.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 89164
; LENGTH: 4283
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_87946C.1
US-10-437-963-89164

```

Query Match	71.2%	Score 17.8;	DB 7;	Length 4283;
Best Local Similarity	90.5%	Pred. No. 8.9e+02;		
Matches	19.	Conservative	0: Mismatches	2: Indels
			0: Gaps	0: Gaps

RESULT 45  
 US-10-719-993-6890  
 ; Sequence 6890, Application US/10719993  
 ; Publication No. US20040265849A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001496  
 ; CURRENT APPLICATION NUMBER: US/10/719,993  
 ; CURRENT FILING DATE: 2003-11-24  
 ; NUMBER OF SEQ ID NOS: 5342  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 6890  
 ; LENGTH: 303172  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature

```

; LOCATION: (1)....(303172)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables
US-10-719-993-6890

```

Query Match	71.2%	Score 17.8;	DB 8;	Length 303172;
Best Local Similarity	90.5%;	Pred. No. 1.8e+03;		
Matches 19: Conservative	0;	Mismatches 2;	Indels 0;	Caps 0;

RESULT 46  
 US-10-424-599-6234  
 ; Sequence 6234, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 6234  
 ; LENGTH: 379  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(379)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_105639C.1  
 ; US-10-424-599-6234

Query Match	70.4%	Score 17.6;	DB 7;	Length 379;
Best Local Similarity	83.3%	Pred. No. 7.1e+02;		
Matches	20:	Conservative	0: Mismatches	4: Indels
				0: Gaps

```

RESULT 47
US-10-425-115-62193/c
; Sequence 62193, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 62193
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_156715C.1
US-10-425-115-62193
Query Match 70.4%; Score 17.6; DB 8; Length 389;

```

```
Best Local Similarity 83.3%; Pred. No. 7.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
   ||| ||||| ||| ||||| |||
Db 372 AAACACTGCATCTCTATGTAATG 349

RESULT 48
US-09-918-995-26833/c
; Sequence 26833, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hvaeq. Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26833
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(461)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26833

Query Match 70.4%; Score 17.6; DB 3; Length 461;
Best Local Similarity 83.3%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 2 AAAAACTGCATCCCAATGTTATG 25
   ||| ||||| ||| ||||| |||
Db 209 AAACACTGCATCAACACTGTTAAG 186

RESULT 49
US-09-925-065A-464930/c
; Sequence 464930, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 464930
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-464930

Query Match 70.4%; Score 17.6; DB 4; Length 519;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;

Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
   ||| ||||| ||| ||||| |||
Db 503 AAATGCTGCATCACCAGATGTTT 480

RESULT 50
US-10-027-632-180366/c
; Sequence 180366, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180366
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-180366

Query Match 70.4%; Score 17.6; DB 5; Length 520;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTAT 24
   ||| ||||| ||| ||||| |||
Db 268 AAATACTGGATCCCAAGATAT 245

Search completed: February 3, 2006, 15:44:02
Job time : 375.556 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-25

Perfect score: 25

Sequence: 1 aaaaactgcattcccaatgttatg 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA.New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22.4	89.6	1400	8	US-11-136-527-6248
2	22.4	89.6	4019	8	US-11-136-527-2152
3	20.4	81.6	1917	7	US-10-793-626-1467
4	20.4	81.6	3276	7	US-10-793-626-4240
5	19.8	79.2	95050	7	US-10-857-780-7
6	19.4	77.6	25	8	US-11-136-527-295557
7	17.6	70.4	201	7	US-10-995-561-55567
8	17.6	70.4	709	7	US-10-750-185-36949
9	17.6	70.4	709	7	US-10-750-623-36949
10	17.6	70.4	720	7	US-10-750-185-48171
11	17.6	70.4	720	7	US-10-750-623-48171
12	17.2	68.8	201	8	US-10-995-561-55566
13	17.2	68.8	201	8	US-11-124-368A-19759
14	17.2	68.8	201	8	US-11-124-368A-19877
15	17.2	68.8	600	7	US-10-750-185-1670
16	17.2	68.8	600	7	US-10-750-623-1670
17	17.2	68.8	855	7	US-10-750-185-41308
18	17.2	68.8	855	7	US-10-750-623-41308
19	17.2	68.8	1130	7	US-10-750-185-52076
20	17.2	68.8	1130	7	US-10-750-623-52076
21	17.2	68.8	79134	8	US-11-124-368A-2924
22	17.2	68.8	218821	8	US-11-121-086-31

Sequence 13327, A	7	US-10-995-561-13327	68.8	244196	17.2	C 23
Sequence 22, Appl	7	US-10-933-025-22	68.8	268685	17.2	C 24
Sequence 827, App	7	US-10-750-185-827	68.0	600	17	C 25
Sequence 827, App	7	US-10-750-623-827	68.0	600	17	C 26
Sequence 73, Appl	8	US-11-121-086-73	68.0	152335	17	C 27
Sequence 13233, A	7	US-10-995-561-13233	68.0	199130	17	C 28
Sequence 3, Appl1	8	US-11-102-978-3	68.0	340000	17	C 29
Sequence 64, Appl	8	US-11-121-086-64	67.2	142605	16.8	C 30
Sequence 63, Appl	8	US-11-121-086-63	67.2	169725	16.8	C 31
Sequence 13215, A	7	US-10-995-561-13215	67.2	241805	16.8	C 32
Sequence 19474, A	7	US-10-750-185-19474	66.4	600	16.6	C 33
Sequence 19474, A	7	US-10-750-623-19474	66.4	600	16.6	C 34
Sequence 24537, A	7	US-10-750-185-24537	66.4	1463	16.6	C 35
Sequence 24537, A	7	US-10-750-623-24537	66.4	1463	16.6	C 36
Sequence 94, Appl	10	US-11-033-764-94	66.4	2161	16.6	C 37
Sequence 29455, A	7	US-10-750-185-29455	66.4	2524	16.6	C 38
Sequence 29455, A	7	US-10-750-623-29455	66.4	2524	16.6	C 39
Sequence 5, Appl1	8	US-11-192-967-5	66.4	5592	16.6	C 40
Sequence 53, Appl1	8	US-11-193-715-5	66.4	5592	16.6	C 41
Sequence 61, Appl	8	US-11-112-908-63	66.4	149111	16.6	C 42
Sequence 61, Appl	8	US-11-112-908-61	66.4	159497	16.6	C 43
Sequence 44, Appl	8	US-11-121-086-44	66.4	167116	16.6	C 44
Sequence 43, Appl	8	US-11-121-086-43	66.4	175416	16.6	C 45
Sequence 20, Appl	8	US-11-112-908-20	66.4	191331	16.6	C 46
Sequence 56163, A	7	US-10-750-185-56163	65.6	1299	16.4	C 47
Sequence 56163, A	7	US-10-750-623-56163	65.6	1299	16.4	C 48
Sequence 2599, Ap	8	US-11-136-527-2599	65.6	2217	16.4	C 49
Sequence 39, Appl	8	US-11-121-086-39	65.6	207835	16.4	C 50
Sequence 40, Appl	8	US-11-121-086-40	65.6	207835	16.4	C 51
Sequence 3061, Ap	7	US-10-750-185-3061	64.8	600	16.2	C 52
Sequence 3061, Ap	7	US-10-750-623-3061	64.8	600	16.2	C 53
Sequence 3367, Ap	8	US-11-128-061-3367	64.8	978	16.2	C 54
Sequence 7009, Ap	8	US-11-128-061-7009	64.8	978	16.2	C 55
Sequence 3367, Ap	8	US-11-128-049-3367	64.8	978	16.2	C 56
Sequence 7009, Ap	8	US-11-128-049-7009	64.8	978	16.2	C 57
Sequence 226, App	5	US-09-978-360A-226	64.8	1283	16.2	C 58
Sequence 1325, A	7	US-10-995-561-1325	64.8	46089	16.2	C 59
Sequence 33, Appl	8	US-11-121-086-33	64.8	154548	16.2	C 60
Sequence 493604, Sequence 62, Appl	25	US-11-121-849-493604	64.0	25	16	C 61
Sequence 48082, A	77	US-10-310-914A-62	64.0	201	16	C 62
Sequence 3068, Ap	7	US-10-995-561-48082	64.0	408	16	C 63
Sequence 1203, Ap	8	US-11-043-752-3068	64.0	408	16	C 64
Sequence 4845, Ap	8	US-11-128-061-1203	64.0	580	16	C 65
Sequence 1203, Ap	8	US-11-128-061-4845	64.0	580	16	C 66
Sequence 4845, Ap	8	US-11-128-049-1203	64.0	580	16	C 67
Sequence 4873, A	8	US-11-128-049-4845	64.0	580	16	C 68
Sequence 40873, A	7	US-10-750-185-40873	64.0	754	16	C 69
Sequence 32644, A	7	US-10-750-623-40873	64.0	754	16	C 70
Sequence 32644, A	7	US-10-750-185-32644	64.0	787	16	C 71
Sequence 51825, A	7	US-10-750-623-51825	64.0	832	16	C 72
Sequence 51825, A	7	US-10-750-185-51825	64.0	832	16	C 73
Sequence 2345, Ap	7	US-10-750-623-51825	64.0	832	16	C 74
Sequence 34, Appl	6	US-10-793-626-2345	64.0	885	16	C 75
Sequence 28120, A	6	US-10-981-334-34	64.0	999	16	C 76
Sequence 100, App	7	US-10-750-185-28120	64.0	1249	16	C 77
Sequence 53359, A	8	US-11-133-345-100	64.0	1557	16	C 78
Sequence 53359, A	7	US-10-750-623-53359	64.0	1665	16	C 79
Sequence 54888, A	7	US-10-750-185-54888	64.0	1669	16	C 80
Sequence 54888, A	7	US-10-750-623-54888	64.0	1669	16	C 81
Sequence 57434, A	7	US-10-750-185-57434	64.0	1726	16	C 82
Sequence 57434, A	7	US-10-750-623-57434	64.0	1726	16	C 83
Sequence 901, App	8	US-11-043-752-901	64.0	2001	16	C 84
Sequence 3469, Ap	7	US-10-793-626-3469	64.0	2489	16	C 85
Sequence 52920, A	7	US-10-750-185-52920	64.0	2670	16	C 86
Sequence 52920, A	7	US-10-750-623-52920	64.0	2670	16	C 87
Sequence 44806, A	7	US-10-750-185-44806	64.0	2824	16	C 88
Sequence 44806, A	7	US-10-750-623-44806	64.0	2824	16	C 89
Sequence 4170, Ap	7	US-10-793-626-4170	64.0	2902	16	C 90
Sequence 4170, Ap	7	US-10-793-626-4170	64.0	2902	16	C 91
Sequence 4120, Ap	7	US-10-793-626-4120	64.0	2929	16	C 92
Sequence 150, App	8	US-11-194-246-150	64.0	3000	16	C 93
			64.0		16	C 94
			64.0		16	C 95

```

c 96 16 64.0 3176 7 US-10-750-185-43645
c 97 16 64.0 3176 7 US-10-750-623-43645
c 98 16 64.0 3350 7 US-10-793-626-4018
c 99 16 64.0 3363 7 US-10-750-185-59149
c 100 16 64.0 3363 7 US-10-750-623-59149
c 101 16 64.0 3660 7 US-10-793-626-4116
c 102 16 64.0 3861 8 US-11-152-892-2
c 103 16 64.0 3911 10 US-11-004-057-5
c 104 16 64.0 4120 8 US-11-136-527-3046
c 105 16 64.0 59590 8 US-11-117-187-187
c 106 16 64.0 98309 8 US-11-124-368A-2921
c 107 16 64.0 151828 8 US-11-117-187-197
c 108 16 64.0 155989 8 US-11-121-086-57
c 109 16 64.0 170925 8 US-11-121-086-35
c 110 16 64.0 178877 8 US-11-121-086-17
c 111 16 64.0 180654 8 US-11-121-086-58
c 112 16 64.0 184000 8 US-11-121-086-37
c 113 16 64.0 645179 7 US-10-995-561-13293
c 114 16 64.0 1691140 8 US-11-091-018-1
c 115 15.8 63.2 6040 7 US-10-240-708-69
c 116 15.8 63.2 35100 8 US-11-127-832-26
c 117 15.8 63.2 162289 8 US-11-121-086-20
c 118 15.8 63.2 1080000 7 US-10-928-446A-1
c 119 15.8 63.2 1080000 7 US-10-928-446A-181
c 120 15.8 63.2 1080000 7 US-10-928-446A-183
c 121 15.8 63.2 1080000 7 US-10-928-446A-185
c 122 15.8 63.2 1080000 7 US-10-928-446A-187
c 123 15.8 63.2 1080000 7 US-10-928-446A-189
c 124 15.8 63.2 1080000 7 US-10-928-446A-191
c 125 15.8 63.2 1080000 7 US-10-928-446A-193
c 126 15.8 63.2 1080000 7 US-10-928-446A-195
c 127 15.8 63.2 1080000 7 US-10-928-446A-197
c 128 15.8 63.2 1080000 7 US-10-928-446A-199
c 129 15.8 63.2 1080000 7 US-10-928-446A-201
c 130 15.6 62.4 25 7 US-10-310-914A-148691
c 131 15.6 62.4 25 8 US-11-121-849-97549
c 132 15.6 62.4 201 7 US-10-995-561-57454
c 133 15.6 62.4 201 7 US-10-995-561-60293
c 134 15.6 62.4 201 7 US-10-995-561-76756
c 135 15.6 62.4 201 7 US-10-995-561-76969
c 136 15.6 62.4 817 7 US-10-750-185-50157
c 137 15.6 62.4 817 7 US-10-750-623-50157
c 138 15.6 62.4 918 8 US-11-074-176-289
c 139 15.6 62.4 1073 7 US-10-750-185-31435
c 140 15.6 62.4 1073 7 US-10-750-623-31435
c 141 15.6 62.4 1107 7 US-10-750-185-59865
c 142 15.6 62.4 1107 7 US-10-750-623-59865
c 143 15.6 62.4 1382 7 US-10-750-185-39793
c 144 15.6 62.4 1382 7 US-10-750-623-39793
c 145 15.6 62.4 1887 7 US-10-793-626-2843
c 146 15.6 62.4 2112 7 US-10-821-234-560
c 147 15.6 62.4 2242 8 US-11-169-041-93
c 148 15.6 62.4 2284 7 US-10-453-372-125
c 149 15.6 62.4 2470 7 US-10-453-372-121
c 150 15.6 62.4 2470 7 US-10-453-372-123
```

ALIGNMENTS

```

RESULT 1
US-11-136-527-6248
; Sequence 6248, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
```

```

; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6248
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6248

Query Match      89.6%; Score 22.4; DB 8; Length 1400;
Best Local Similarity 95.8%; Pred. No. 0.31; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 1;

QY      2 AAAAAGTGCATCCCAATGTTATG 25
      |||
Db      1296 AAAAAGTGCATCCCAATGTTATG 1319
      |||

RESULT 2
US-11-136-527-2152
; Sequence 2152, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2152
; LENGTH: 4019
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2152

Query Match      89.6%; Score 22.4; DB 8; Length 4019;
Best Local Similarity 95.8%; Pred. No. 0.38; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 1;

QY      2 AAAAAGTGCATCCCAATGTTATG 25
      |||
Db      3915 AAAAAGTGCATCCCAATGTTATG 3938
      |||

RESULT 3
US-10-793-626-1467
; Sequence 1467, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1467
; LENGTH: 1917
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1467

Query Match      81.6%; Score 20.4; DB 7; Length 1917;
Best Local Similarity 95.5%; Pred. No. 2.9; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1;

US-10-793-626-1467
```

```
Qy 3 AAAAAGTCATCCCAATGTTAT 24
Db 1441 AAAAAGTCATCCCTCAATGTTAT 1462

RESULT 4
US-10-793-626-4240
; Sequence 4240, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,358
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4240
; LENGTH: 3276
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4240

Query Match : 81.6%; Score 20.4; DB 7; Length 3276;
Best Local Similarity 95.5%; Pred. No. 3.2;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAAAAGTCATCCCAATGTTAT 24
Db 1193 AAAAAGTCATCCCTCAATGTTAT 1214

RESULT 5
US-10-857-780-7
; Sequence 7, Application US/10857780
; Publication No. US2005027043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENEALAND, RIKARD HENRY
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 95050
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (18543)..(18543)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28737)..(28737)

OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (45763)..(45763)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (51178)..(51178)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (51180)..(51180)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (62241)..(62241)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (62541)..(62541)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (64042)..(64042)
OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-7

Query Match : 79.2%; Score 19.8; DB 7; Length 95050;
Best Local Similarity 91.3%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAGTCATCCCAATGTTA 23
Db 29898 AAACAACTGCAACCCCAATGTTA 29920

RESULT 6
US-11-136-527-295557
; Sequence 295557, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 295557
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-295557

Query Match : 77.6%; Score 19.4; DB 8; Length 25;
Best Local Similarity 95.2%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 AACTGCATCCCAATGTTATG 25
Db 1 AACTGCATCCCAATGTTATG 21

RESULT 7
US-10-995-561-55567/c
; Sequence 55567, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
```



; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 48171  
; LENGTH: 720  
; TYPE: DNA  
; ORGANISM: Bovine 19866880932834  
US-10-750-623-48171

Query Match 70.4%; Score 17.6; DB 7; Length 720;  
Best Local Similarity 83.3%; Pred. No. 50;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTTAT 24  
Db 349 ATAAAACTGAATCCACAATGTTCT 372

RESULT 12  
US-10-995-561-55566/c  
; Sequence 55566, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 55566  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-55566

Query Match 68.8%; Score 17.2; DB 7; Length 201;  
Best Local Similarity 79.2%; Pred. No. 60;  
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAACTGCATCCCAATGTTAT 24  
Db 102 AMCAAAATGCATCCCACTGTCAT 79

RESULT 13  
US-11-124-368A-19759  
; Sequence 19759, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 19759  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-19759

Query Match 68.8%; Score 17.2; DB 8; Length 201;

Best Local Similarity 86.4%; Pred. No. 60;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 AAAAACTGCATCCCAATGTTA 23  
Db 101 AAAAACTTAATCCCAATGTGA 122

## RESULT 14

US-11-124-368A-19877  
; Sequence 19877, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 19877  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-19877

Query Match 68.8%; Score 17.2; DB 8; Length 201;  
Best Local Similarity 86.4%; Pred. No. 60;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCAATGTTA 23  
Db 132 AAAAACTTAATCCCAATGTGA 153

## RESULT 15

US-10-750-185-1670  
; Sequence 1670, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 1670  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT19761  
US-10-750-185-1670

Query Match 68.8%; Score 17.2; DB 7; Length 600;  
Best Local Similarity 79.2%; Pred. No. 75;  
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCAATGTTATG 25



Db 437 AMCACTGGACCCCAATTATG 460

RESULT 16  
US-10-750-623-1670  
; Sequence 1670, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1670  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MWBT19761  
US-10-750-623-1670

Query Match 68.8%; Score 17.2; DB 7; Length 600;  
Best Local Similarity 79.2%; Pred. No. 75;  
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 AAAAACTGCATCCCCCAATGTTATG 25  
Db 437 AMCACTGGACCCCAATTATG 460

RESULT 17  
US-10-750-185-41308/c  
; Sequence 41308, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41308  
; LENGTH: 855  
; TYPE: DNA  
; ORGANISM: Bovine 19866881259141  
US-10-750-185-41308

Query Match 68.8%; Score 17.2; DB 7; Length 855;  
Best Local Similarity 86.4%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAACTGCATCCCCCAATGTT 22  
Db 411 AAAAAATTCATCCCCCAATGTT 390

RESULT 18  
US-10-750-623-41308/c  
; Sequence 41308, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41308  
; LENGTH: 855  
; TYPE: DNA  
; ORGANISM: Bovine 19866881259141  
US-10-750-623-41308

Query Match 68.8%; Score 17.2; DB 7; Length 855;  
Best Local Similarity 86.4%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAACTGCATCCCCCAATGTT 22  
Db 411 AAAAAATTCATCCCCCAATGTT 390

RESULT 19  
US-10-750-185-52076/c  
; Sequence 52076, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 52076  
; LENGTH: 1130  
; TYPE: DNA  
; ORGANISM: Bovine 19866880733723  
US-10-750-185-52076

Query Match 68.8%; Score 17.2; DB 7; Length 1130;  
Best Local Similarity 86.4%; Pred. No. 84;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 AAAAACTGCATCCCCCAATGTTA 23  
Db 85 AAAAAATTCATCCCCCAATTTTA 64

RESULT 20  
US-10-750-623-52076/c

```
; Sequence 52076, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52076
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Bovine 19866880733723
; US-10-750-623-52076

Query Match      68.8%; Score 17.2; DB 7; Length 1130;
Best Local Similarity 86.4%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCAATGTTA 23
   ||||| ||| ||||| |||
Db 85 AAAAAATGCCCTCCCAATTTA 64

RESULT 21
US-11-124-368A-2924
; Sequence 2924, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2924
; LENGTH: 79134
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-124-368A-2924

Query Match      68.8%; Score 17.2; DB 8; Length 79134;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAACTGCATCCCAATGTTA 23
   ||||| ||| ||||| |||
Db 50290 AAAAACTAATCCCAATGTGA 50311

RESULT 22
US-11-121-086-31/c
; Sequence 31, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
; LENGTH: 218821
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (106949)..(106949)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (110322)..(110324)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (115133)..(115133)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (131300)..(131300)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139059)..(139158)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157740)..(157740)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157777)..(157777)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157900)..(157900)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157919)..(157919)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157926)..(157926)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158094)..(158094)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158138)..(158138)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158193)..(158195)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158241)..(158242)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158259)..(158259)
; OTHER INFORMATION: a, c, g, t, unknown or other
```

```
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158278)..(158278)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158295)..(158295)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158740)..(158839)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158929)..(158929)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (163528)..(163530)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (163550)..(163550)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (163785)..(163785)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (164000)..(164000)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (164047)..(164047)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (164084)..(164084)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (167233)..(167233)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (167236)..(167236)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (167238)..(167238)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170247)..(170247)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170249)..(170250)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170252)..(170253)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170259)..(170259)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170262)..(170263)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
```

```
; NAME/KEY: modified_base
; LOCATION: (170266)..(170266)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (174470)..(174470)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (174472)..(174472)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (174474)..(174474)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (179059)..(179060)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (179064)..(179064)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (197001)..(197001)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (197004)..(197005)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (197007)..(197007)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (200349)..(200349)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (200351)..(200351)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (212425)..(212426)
; OTHER INFORMATION: a, c, g, t, unknown or other
; US-11-121-086-31
```

```
Query Match      68.8%; Score 17.2; DB 8; Length 218821;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1  AAAAACTGCATCCCAATGTT 22
          |||||
Db      17443  AAAGACTGCATACCGAATGTT 17422
```

```
RESULT 23
US-10-995-561-13327/c
; Sequence 13327, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13327
```

```
; LENGTH: 244196
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(244196)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13327

Query Match      68.8%; Score 17.2; DB 7; Length 244196;
Best Local Similarity 79.2%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 1; Mismatches 4;

QY 1 AAAAACTGCATCCCAATGTTAT 24
    |||||
Db 28164 AMCAATGCATCCCACTGTAT 28141
    |||||

RESULT 24
US-10-933-025-22/c
; Sequence 22, Application US/10933025
; Publication No. US20050265987A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN, STEVEN
; APPLICANT: HEMMERICH, STEFAN
; APPLICANT: TOMITA, MEGUMI
; TITLE OF INVENTION: Sulfotransferases and methods of use
; FILE REFERENCE: UCAL-230CON
; CURRENT APPLICATION NUMBER: US/10/933,025
; CURRENT FILING DATE: 2004-09-01
; PRIOR APPLICATION NUMBER: 10/025,966
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/258,577
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 60/267,831
; PRIOR FILING DATE: 2001-09-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 268685
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(268685)
; OTHER INFORMATION: n = A,T,C or G
US-10-933-025-22

Query Match      68.8%; Score 17.2; DB 7; Length 268685;
Best Local Similarity 86.4%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3;

QY 3 AAAAAGTCATCCCAATGTTAT 24
    |||||
Db 265891 AATACTGCATCCGCAATGTAT 265870
    |||||

RESULT 25
US-10-750-185-827
; Sequence 827, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
```

```
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 827
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT09456
US-10-750-185-827

Query Match      68.0%; Score 17; DB 7; Length 600;
Best Local Similarity 80.0%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    |||||
Db 176 AAAAAATCATCCCTTAAGTTATG 200
    |||||

RESULT 26
US-10-750-623-827
; Sequence 827, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 827
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT09456
US-10-750-623-827

Query Match      68.0%; Score 17; DB 7; Length 600;
Best Local Similarity 80.0%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
    |||||
Db 176 AAAAAATCATCCCTTAAGTTATG 200
    |||||

RESULT 27
US-11-121-086-73
; Sequence 73, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 73
; LENGTH: 152335
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-11-121-086-73

Query Match      68.0%; Score 17; DB 8; Length 152335;
Best Local Similarity 80.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 150782 ACAAATCGAATCCCAATGTTATG 150806

RESULT 28
US-10-995-561-13233
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1)..(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      68.0%; Score 17; DB 7; Length 199130;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 56718 AACAATCTGCATCCCAATGTTATG 56742

RESULT 29
US-11-102-978-3
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537-1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948)..(57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80006)..(81089)

; OTHER INFORMATION: Gene VDACP; voltage-dependent anion channel isoform 2 pseudogene
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308)..(167438)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732)..(216833)
; OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match      68.0%; Score 17; DB 8; Length 340000;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGCATCCCAATGTTATG 25
   ||||| ||||| ||||| ||||| |||||
Db 182887 AAAAAAATGCTTACCAATGGTGTG 182911

RESULT 30
US-11-121-086-64/c
; Sequence 64, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 64
; LENGTH: 142605
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-64

Query Match      67.2%; Score 16.8; DB 8; Length 142605;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACTGCATCCCAATGTTA 23
   ||||| ||||| ||||| ||||| |||||
Db 26350 AAACTCCATCCCAATGTGA 26331

RESULT 31
US-11-121-086-63/c
; Sequence 63, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 63
; LENGTH: 169725
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-63

Query Match      67.2%; Score 16.8; DB 8; Length 169725;
Best Local Similarity 90.0%; Pred. No. 3.3e+02;
```

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AACTGCATCCCAATGTTA 23  
Db 168192 AACTGCATCCCAATGTTA 168173

## RESULT 32

US-10-995-561-13215  
; Sequence 13215, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; PRIOR FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13215  
; LENGTH: 241805  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(241805)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-995-561-13215

Query Match 67.2%; Score 16.8; DB 7; Length 241805;  
Best Local Similarity 90.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 AACTGCATCCCAATGTTAT 24  
Db 78480 AACTGCACCCCAAGTTAT 78499

## RESULT 33

US-10-750-185-19474/c  
; Sequence 19474, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19474  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT04310  
US-10-750-185-19474

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAGCGCATCCCAATGTTAT 24  
Db 568 AAAAAGCGCATCCCAATGTTAT 546

## RESULT 34

US-10-750-623-19474/c  
; Sequence 19474, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19474  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT04310  
US-10-750-623-19474

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAGCGCATCCCAATGTTAT 24  
Db 568 AAAAAGCGCATCCCAATGTTAT 546

## RESULT 35

US-10-750-185-24537/c  
; Sequence 24537, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24537  
; LENGTH: 1463  
; TYPE: DNA  
; ORGANISM: Bovine 19866880525178  
US-10-750-185-24537

Query Match 66.4%; Score 16.6; DB 7; Length 1463;  
Best Local Similarity 82.6%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAGCGCATCCCAATGTTAT 24  
Db 1070 AAAAAGCGCATCCCAATGTTAT 1048

## RESULT 36





```

Query Match      66.4%; Score 16.6; DB 8; Length 5592;
Best Local Similarity 82.6%; Pred. NO. 2.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0

Qy 2 AAAAAGCTGCATCCCAATGTTAT 24
    |||||
Db 1700 AATAATTGCTCCCAATGTTAT 1678

```

```
Query Match      66.4%; Score 16.6; DB 8; Length 159497;
Best Local Similarity 82.6%; Pred.No.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

RESULT 44  
US-11-121-086-44  
; Sequence 44, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 44  
; LENGTH: 167116  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-44

Query Match 66.4%; Score 16.6; DB 8; Length 167116;  
Best Local Similarity 82.6%; Pred. No. 4.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTGCATCCCAATGTTA 23  
DB 1917 AGAAATCTGTATCCCAATGTTA 1939

RESULT 45  
US-11-121-086-43  
; Sequence 43, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 43  
; LENGTH: 175416  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-43

Query Match 66.4%; Score 16.6; DB 8; Length 175416;  
Best Local Similarity 82.6%; Pred. No. 4.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTGCATCCCAATGTTA 23  
DB 168998 AGAAATCTGTATCCCAATGTTA 169020

RESULT 46  
US-11-112-908-20/c  
; Sequence 20, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23

; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 20  
; LENGTH: 191331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-20

Query Match 66.4%; Score 16.6; DB 8; Length 191331;  
Best Local Similarity 82.6%; Pred. No. 4.2e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAACTGCATCCCAATGTTAT 24  
DB 4946 AAAAAATTTCTCCCAATATTAT 4924

RESULT 47  
US-10-750-185-56163/c  
; Sequence 56163, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56163  
; LENGTH: 1299  
; TYPE: DNA  
; ORGANISM: Bovine  
US-10-750-185-56163

Query Match 65.6%; Score 16.4; DB 7; Length 1299;  
Best Local Similarity 94.4%; Pred. No. 2.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAACTGCATCCCAAA 18  
DB 164 AAAAACTGAATCCCAAA 147

RESULT 48  
US-10-750-623-56163/c  
; Sequence 56163, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AAAGTGCATCCCCAATGT 21  
||||| |||||||  
Db 134276 AAAGTGCATCCCCAATGT 134293

Search completed: February 3, 2006, 16:20:17  
Job time : 353.111 secs

;; CURRENT FILING DATE: 2003-12-31  
;; PRIOR APPLICATION NUMBER: US 60/437,482  
;; PRIOR FILING DATE: 2002-12-31  
;; NUMBER OF SEQ ID NOS: 64922  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 56163  
;; LENGTH: 1299  
;; TYPE: DNA  
;; ORGANISM: Bovine 19866880487873  
US-10-750-623-56163

Query Match 65.6%; Score 16.4; DB 7; Length 1299;  
Best Local Similarity 94.4%; Pred. No. 2.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAGTGCATCCCCAA 18  
||||| |||||||  
Db 164 AAAAAGTGAATCCCCAA 147

RESULT 49  
US-11-136-527-2599/c  
; Sequence 2599, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2599  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-2599

Query Match 65.6%; Score 16.4; DB 8; Length 2217;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 CTCGATCCCCCAATGTTAT 24  
||||| |||||||  
Db 226 CTCGATCCCCCAATGTTAT 209

RESULT 50  
US-11-121-086-39  
; Sequence 39, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 39  
; LENGTH: 207835  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-39

Query Match 65.6%; Score 16.4; DB 8; Length 207835;  
Best Local Similarity 94.4%; Pred. No. 5.1e+02;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)

1646.682 Million cell updates/sec

Title: US-10-719-900-30

Perfect score: 25

Sequence: 1 aaaaactggaatctcagctgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_ste.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_hcg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	5223	9	AK173165 Mus muscu
C 2	25	100.0	195988	9	AC092481 Mus muscu
C 3	25	100.0	213144	9	AC092480 Mus muscu
4	25	100.0	214356	14	AC154650 Mus muscu
5	25	100.0	243637	14	AC097150 Rattus no
6	25	100.0	248916	14	AC137217 Rattus no
C 7	20.2	80.8	25837	14	AC087198 Homo sapi
C 8	20.2	80.8	30005	8	AL596124 Human DNA
C 9	20.2	80.8	180569	8	AL354949 Human DNA
C 10	20.2	80.8	193379	14	AC156206 Bos tauru
C 11	20.2	80.8	194019	14	AC053464 Homo sapi
12	20.2	80.8	208202	14	AC013446 Homo sapi
C 13	20.2	80.8	231342	14	AC152756 Bos tauru
14	20	80.0	300	6	BD212075 Novel hum
15	20	80.0	758	6	BD215003 Novel hum
16	20	80.0	2272	8	AB168750 Macaca fa
17	20	80.0	4810	8	AB037818 Homo sapi
18	20	80.0	5174	6	AX405807 Sequence

19	20	80.0	5183	8	AF288480 Homo sapi
20	20	80.0	11127	6	CS033671 Sequence
21	20	80.0	11127	6	CS042623 Sequence
22	20	80.0	11127	8	AF219946 Homo sapi
C 23	20	80.0	180635	8	AL360169 Human DNA
24	20	80.0	299920	14	AC108611 Rattus no
25	19.8	79.2	134514	14	AC025192 Homo sapi
26	19.8	79.2	162560	14	AC069005 Homo sapi
27	19.8	79.2	167878	8	AC103719 Homo sapi
28	19.8	79.2	189662	8	AC015468 Homo sapi
C 29	19.8	79.2	203690	14	AC087221 Homo sapi
C 30	19.8	79.2	225793	14	AC156131 Bos tauru
C 31	19.8	79.2	227115	14	AC152300 Bos tauru
C 32	19.8	79.2	234429	8	AF311103 Homo sapi
C 33	19.8	77.6	168813	8	AC013717 Homo sapi
34	19.2	76.8	54666	8	AC073487 Homo sapi
C 35	19.2	76.8	62880	14	AC026926 Homo sapi
36	19.2	76.8	99176	5	BX255893 Zebrafish
C 37	19.2	76.8	105736	14	AC090678 Homo sapi
C 38	19.2	76.8	151328	14	BX936340 Dantio rer
C 39	19.2	76.8	152222	9	AC158544 Mus muscu
C 40	19.2	76.8	152267	8	AC104037 Homo sapi
C 41	19.2	76.8	154170	9	AL671979 Mouse DNA
C 42	19.2	76.8	155268	9	AC132332 Mus muscu
C 43	19.2	76.8	155669	14	AF307158 Homo sapi
C 44	19.2	76.8	162528	9	AC122463 Mus muscu
45	19.2	76.8	164792	14	AC119551 Rattus no
C 46	19.2	76.8	180462	9	AC132235 Mus muscu
47	19.2	76.8	192020	9	AC130713 Mus muscu
C 48	19.2	76.8	197585	9	AC114552 Mus muscu
C 49	19.2	76.8	197966	14	AC118371 Rattus no
C 50	19.2	76.8	213357	14	AC166021 Oryctolag
C 51	19.2	76.8	213599	14	AC140764 Rattus no
C 52	19.2	76.8	222112	14	AC163602 Bos tauru
C 53	19.2	76.8	228741	9	AC134902 Mus muscu
54	19.2	76.8	234998	14	AC108266 Rattus no
C 55	19.2	76.8	246860	14	AC130169 Rattus no
C 56	19.2	76.8	248151	14	AC157338 Bos tauru
57	19.2	76.8	253819	14	AC095474 Rattus no
58	19.2	76.8	286491	14	AC096988 Rattus no
C 59	18.8	75.2	39852	8	AC006275 Homo sapi
C 60	18.8	75.2	55266	8	AY192161 Homo sapi
C 61	18.8	75.2	75782	8	AC092068 Homo sapi
C 62	18.8	75.2	109377	8	AL355994 Human DNA
C 63	18.8	75.2	110000	1	CP000107 05 Continuation (6 of
64	18.8	75.2	110000	1	CP000058 56 Continuation (57 o
C 65	18.8	75.2	110000	14	BX005127 2 Continuation (3 of
66	18.8	75.2	157470	14	AC137446 Rattus no
C 67	18.8	75.2	161206	14	AC114249 Rattus no
C 68	18.8	75.2	163599	8	HS078108 Human DNA
69	18.8	75.2	171097	9	AC112932 Mus muscu
70	18.8	75.2	175896	14	AC165317 Mus muscu
71	18.8	75.2	177878	14	AC010893 Homo sapi
72	18.8	75.2	179154	14	AC084862 Papio anu
73	18.8	75.2	186608	8	AL353734 Human DNA
74	18.8	75.2	187079	14	AC083911 Rattus no
C 75	18.8	75.2	189994	14	AC124882 Rattus no
76	18.8	75.2	203566	14	AC142173 Rattus no
C 77	18.8	75.2	204483	9	AL671990 Mouse DNA
78	18.8	75.2	207051	14	AC110030 Mus muscu
C 79	18.8	75.2	229303	14	AC113844 Rattus no
C 80	18.8	75.2	241154	14	AC160800 Bos tauru
C 81	18.8	75.2	254778	14	AC126532 Rattus no
82	18.8	75.2	271532	14	AC122958 Rattus no
C 83	18.8	75.2	271932	14	AC109948 Rattus no
84	18.8	75.2	306769	14	AL356105 Homo sapi
85	18.6	74.4	600	10	BV410946 S219P6527
86	18.6	74.4	718	10	BV549133 S219P6590
87	18.6	74.4	779	9	RNS1P1 X07284 Rat gene fo
C 88	18.6	74.4	779	10	BV535337 G591P6288
C 89	18.6	74.4	794	10	BV529962 G591P6170
C 90	18.6	74.4	830	10	BV542893 s2m51h08
C 91	18.6	74.4	842	15	DCU52847 Daucus caro

C 92	18.6	74.4	850	15	DCU52848	U52848 Daucus caro
C 93	18.6	74.4	1000	10	BV545071	BV545071 sxe63D03.
C 94	18.6	74.4	1047	4	AB198070	AB198070 Bos tauru
C 95	18.6	74.4	1364	8	BSM807436	BSM807436 Homo sapi
C 96	18.6	74.4	1388	8	AK126147	AK126147 Homo sapi
C 97	18.6	74.4	2066	6	AX713787	AX713787 Sequence
C 98	18.6	74.4	2066	8	AK055576	AK055576 Homo sapi
C 99	18.6	74.4	4527	15	PANMTC01	M28703 Podospora a
C 100	18.6	74.4	29227	2	AF000192	AF000192 Caenorhab
C 101	18.6	74.4	30223	1	AE001576	AE001576 Borrelia
C 102	18.6	74.4	30316	8	AC124219	AC124219 Pan trogl
C 103	18.6	74.4	40179	8	AC004036	AC004036 Homo sapi
C 104	18.6	74.4	44788	8	AL133267	AL133267 Homo sapi
C 105	18.6	74.4	45878	15	AB016883	AB016883 Arabidops
C 106	18.6	74.4	52371	1	AE001584	AE001584 Borrelia
C 107	18.6	74.4	54398	14	AC068224	AC068224 Homo sapi
C 108	18.6	74.4	63771	8	AL159045	AL159045 Human DNA
C 109	18.6	74.4	68581	8	AP000563	AP000563 Homo sapi
C 110	18.6	74.4	80582	9	AL953839	AL953839 Mouse DNA
C 111	18.6	74.4	82789	14	AC015629	AC015629 Homo sapi
C 112	18.6	74.4	95477	8	AC007076	AC007076 Homo sapi
C 113	18.6	74.4	97777	14	AC165650	AC165650 Bos tauru
C 114	18.6	74.4	99314	14	AC084061	AC084061 Bos muscu
C 115	18.6	74.4	100000	8	AP000211	AP000211 Homo sapi
C 116	18.6	74.4	100314	15	MTPACG	XS5026 Podospora a
C 117	18.6	74.4	100348	8	AC091860	AC091860 Homo sapi
C 118	18.6	74.4	100960	14	AC166764	AC166764 Bos tauru
C 119	18.6	74.4	102649	14	AC134715	AC134715 Rattus no
C 120	18.6	74.4	110000	1	AE006470	Continuation (15 o
C 121	18.6	74.4	110000	14	BX324168	BX324168 Homo sapi
C 122	18.6	74.4	110000	14	TANN2_00	CR940348 Theileria
C 123	18.6	74.4	122001	8	AC007215	AC007215 Homo sapi
C 124	18.6	74.4	125066	8	AC074286	AC074286 Homo sapi
C 125	18.6	74.4	126852	8	AC010478	AC010478 Homo sapi
C 126	18.6	74.4	127675	8	AC146221	AC146221 Pan trogl
C 127	18.6	74.4	130158	14	AC166075	AC166075 Mus muscu
C 128	18.6	74.4	132890	14	AC163197	AC163197 Bos tauru
C 129	18.6	74.4	135146	14	AC161719	AC161719 Loxodonta
C 130	18.6	74.4	137246	8	HS49J10	284572 Human DNA s
C 131	18.6	74.4	139719	14	AC141439	AC141439 Homo sapi
C 132	18.6	74.4	139744	5	BX957315	BX957315 Zebrafish
C 133	18.6	74.4	142948	8	AC147025	AC147025 Pan trogl
C 134	18.6	74.4	146259	8	AL137013	AL137013 Human DNA
C 135	18.6	74.4	149965	8	AC007718	AC007718 Homo sapi
C 136	18.6	74.4	152311	14	AC011571	AC011571 Homo sapi
C 137	18.6	74.4	160606	8	AC093517	AC093517 Homo sapi
C 138	18.6	74.4	161781	14	CR376800	CR376800 Danio rer
C 139	18.6	74.4	162377	8	AL683870	AL683870 Human DNA
C 140	18.6	74.4	163046	9	AC129585	AC129585 Mus muscu
C 141	18.6	74.4	164294	14	AC148910	AC148910 Otolemur
C 142	18.6	74.4	166093	9	AC119971	AC119971 Mus muscu
C 143	18.6	74.4	168487	8	AL138820	AL138820 Human DNA
C 144	18.6	74.4	168799	8	AC009531	AC009531 Homo sapi
C 145	18.6	74.4	169786	8	BS000112	BS000112 Homo sapi
C 146	18.6	74.4	170654	8	AC073901	AC073901 Homo sapi
C 147	18.6	74.4	173021	8	AC146376	AC146376 Pan trogl
C 148	18.6	74.4	176277	14	AC025941	AC025941 Homo sapi
C 149	18.6	74.4	179906	14	AC119339	AC119339 Rattus no
C 150	18.6	74.4	181446	14	CR954266	CR954266 Danio rer

ALIGNMENTS

RESULT 1	AK173165	AK173165	5223 bp	mRNA	linear	ROD 28-JUL-2004
LOCUS	Mus musculus	Mus musculus				
DEFINITION	Mus musculus mRNA for mKIAA1397 protein.					
ACCESSION	AK173165					
VERSION	AK173165.1 GI:50510914					
KEYWORDS	FLI_CDNA.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					

REFERENCE	AUTHORS										
Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Koseki,H., Hiraoka,S., Saga,Y., Seino,S., Nishimura,M., Kaisho,T., Hoshino,K., Kitamura,H., Nagase,T., Ohara,O. and Koga,H.											
Prediction of the Coding Sequences of Mouse Homologues of KIAA											
Gene: IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries											
DNA Res. 11, 205-218 (2004)											
2 (bases 1 to 5223)											
Okazaki,N., Kikuno,R.F., Nagase,T., Ohara,O. and Koga,H.											
Direct Submission											
Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)											
The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing: Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.											
FEATURES											
Location/Qualifiers											
1..5223											
/organism="Mus musculus"											
/mol_type="mRNA"											
/db_xref="taxon:10090"											
/clone="mbg12832"											
/tissue_type="brain"											
/dev_stage="adult"											
note="vector:modified pBC SK+"											
1..5223											
/gene="mKIAA1397"											
<623..2833											
/gene="mKIAA1397"											
/note="CDS is predicted by in silico analysis. Start codon is not identified."											
/codon_start=1											
/evidence=not_experimental											
/product="mKIAA1397 protein"											
/protein_id="BAD32443.1"											
/db_xref="GI:50510915"											
/translation="LDSRSVYTKKKNKPTPNHKNKDKIKELGLKFLSTFKVKYFKIE SMYAAVEHGVLCSDNLCUSKGRVPSKPKVRRRIYEGNLAIGRGGVVG TFTSHCRDRSTPTQFNFNLRHNSVLRWNEPVYKLTADGGIFVMIQYEGR WSELVNDRGAQVSDFTWSDGTQALISYRDGFVLVSGVSGQRHWSINLESQITCG IWTDDQVLFTGADGOVIMDCRGLAHVLLHESDGLSWMYPIFLVEDSSES TDSDYSPPQDPAAYPIPVONTKPLLTVSFTSGDISLMNNVDDLSPVIRSGLKEV AQWCTQGLLAVAGNEQAQSELNPGNLLKSAMVKFYNVRGEHIFLDTLVORPIL ICWHRDRSRLWASGPAIVYVRHVSLLQLCCQALASTREDKDVNKTLPRLIC SYLSTAFPTTKPIPDNNRDFVSPSAGNERLHCTMKRTDDPFVGGPCYTLLE YLGLVAFILKGRISKLKPEFIMDPRTDSKSDIYGNLSLSTVDSNCSDSDIEL SDDWAAKSPKISRSSKPKLPRISEARKSPKLPRAAQEISRSRPLRPMKSPGMS LTRREFPEDITQHNLYLAQVTSNMGTKFVGLAAFLPTNLGACKDSLIVTVLEI NKSHNLNQWFAKFAENEFLRQRTPTVCLLDDSSHLSLLEKQLCSLALFLBLSCVFL"											
ORIGIN											
Query Match 100.0%; Score 25; DB 9; Length 5223;											
Best Local Similarity 100.0%; Pred. No. 0.47;											
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;											
QY	1	AAAAA								CTGGAATCTCAGGCTGAGA	25
Db	381	AAAAA								CTGGAATCTCAGGCTGAGA	405
RESULT 2											
AC092481/c	195988 bp DNA linear ROD 15-NOV-2002										
LOCUS	Mus musculus strain C57BL/6J chromosome 17 clone rp23-456p8,										
DEFINITION	complete sequence.										
ACCESSION	AC092481										







Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, V., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joliviet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Loulsegad, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwokeleleh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Pal, S., Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.-L., Puazo, M., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Streinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

Direct Submission

2 (bases 1 to 243637)

Worley, K.C.

## REFERENCE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

## COMMENT

Submitted (11-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 243637)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:25008195. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

----- Center project name: GIEL

Center clone name: CH230-21612

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 223909 bases at least Q40

Consensus quality: 228092 bases at least Q30

Consensus quality: 231096 bases at least Q20

Estimated insert size: 238161; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 243637: contig of 243637 bp in length.

## FEATURES

source

1. 243637

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-21612"

1. 1443

/note="wgs end extension"

clone\_end:Sp6"

complement(2266..2885)

/note="clone boundary"

clone\_end:Sp6"

site:EcoRI

end\_sequence:B2097561"

167887..170883

/note="wgs contig"

complement(239558..240294)

/note="clone boundary"

clone\_end:T7"

site:EcoRI

end\_sequence:B2097560"

241686..243637

/note="wgs end extension"

clone\_end:T7"

## ORIGIN

Query Match 100.0%; Score 25; DB 14; Length 243637;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTGGATCTCAGGCTGAGA 25

|||||

Db 144610 AAAAACTGGATCTCAGGCTGAGA 144634

|||||

RESULT 6

AC137217

LOCUS

DEFINITION

AC137217.1 GI:25073111

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus clone CH230-unknown, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 4 unordered pieces.

AC137217

AC137217.1

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 248916)

Muzny, D., Marie, J., Metzker, M., Lee, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Danson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nsamenang,A., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pastekernak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

# TITLE JOURNAL

## REFERENCE

## AUTHORS JOURNAL

## COMMENT

2 (bases 1 to 248916)

Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: K2NJ

Center clone name: CH230-unknown  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 221283 bases at least Q40  
Consensus quality: 226132 bases at least Q30  
Consensus quality: 229684 bases at least Q20  
Estimated insert size: 231897; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 242098: contig of 242098 bp in length  
\* 242099 242198: gap of unknown length  
\* 242199 243368: contig of 1170 bp in length  
\* 243369 243468: gap of unknown length  
\* 243469 244799: contig of 1331 bp in length  
\* 244800 244899: gap of unknown length  
\* 244900 248916: contig of 4017 bp in length.  
\* 244900 248916: contig of 4017 bp in length.

## FEATURES source

1. 248916  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-unknown"

## misc\_feature

1..2515  
/note="wgs\_contig"

## misc\_feature

121167..123058  
/note="wgs\_contig"

## misc\_feature

218817..220508  
/note="wgs\_contig"

## gap

242099..242198  
/estimated\_length=unknown

## gap

243369..243468  
/estimated\_length=unknown

## gap

244800..244899  
/estimated\_length=unknown

## ORIGIN

Query Match 100.0%; Score 25; DB 14; Length 248916;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25

Db 150938 AAAAACTGGAATCTCAGGCTGAGA 150962

## RESULT 7

## AC087198/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## 1 (bases 1 to 25837)

## Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## Homo sapiens chromosome 8, clone RP11-5803

## Unpublished

## 2 (bases 1 to 25837)

AC087198 25837 bp DNA linear HTG 13-DEC-2000  
Homo sapiens chromosome 8 clone RP11-5803 map 8, LOW-PASS SEQUENCE  
SAMPLING.

AC087198 GI:11693380

HTG; HTGS PHASE0.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 25837)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 8, clone RP11-5803

Unpublished

2 (bases 1 to 25837)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,  
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagoos,B., Heaford,A., Horton,A., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,  
 Lehoczyk,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,  
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
 Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
 Zemek,L., Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

Submitted (13-DEC-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L11599

Center clone name: 58\_O3

\* NOTE: This record contains 32 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 692: contig of 692 bp in length  
 \* 693 792: gap of 100 bp  
 \* 793 1495: contig of 703 bp in length  
 \* 1496 1595: gap of 100 bp  
 \* 1596 2282: contig of 687 bp in length  
 \* 2283 2382: gap of 100 bp  
 \* 2383 3095: contig of 713 bp in length  
 \* 3096 3195: gap of 100 bp  
 \* 3196 3896: contig of 701 bp in length  
 \* 3897 3996: gap of 100 bp  
 \* 3997 4695: contig of 699 bp in length  
 \* 4696 4795: gap of 100 bp  
 \* 4796 5226: contig of 731 bp in length  
 \* 5227 5626: gap of 100 bp  
 \* 5627 6339: contig of 713 bp in length  
 \* 6340 6440: gap of 100 bp  
 \* 6440 7142: contig of 703 bp in length  
 \* 7143 7242: gap of 100 bp  
 \* 7243 7933: contig of 691 bp in length  
 \* 7934 8033: gap of 100 bp  
 \* 8034 8749: contig of 716 bp in length  
 \* 8750 8849: gap of 100 bp  
 \* 8850 9560: contig of 711 bp in length  
 \* 9561 9660: gap of 100 bp  
 \* 9661 10365: contig of 705 bp in length  
 \* 10366 10465: gap of 100 bp

\* 10466 11186: contig of 721 bp in length  
 \* 11187 11286: gap of 100 bp  
 \* 11287 11999: contig of 713 bp in length  
 \* 12000 12099: gap of 100 bp  
 \* 12099 12809: contig of 710 bp in length  
 \* 12810 12909: gap of 100 bp  
 \* 12910 13608: contig of 699 bp in length  
 \* 13609 13708: gap of 100 bp  
 \* 13709 14404: contig of 696 bp in length  
 \* 14405 14504: gap of 100 bp  
 \* 14505 15224: contig of 720 bp in length  
 \* 15225 15324: gap of 100 bp  
 \* 15325 16048: contig of 724 bp in length  
 \* 16049 16148: gap of 100 bp  
 \* 16149 16864: contig of 716 bp in length  
 \* 16865 16964: gap of 100 bp  
 \* 16965 17677: contig of 713 bp in length  
 \* 17678 17777: gap of 100 bp  
 \* 17778 18494: contig of 717 bp in length  
 \* 18495 18594: gap of 100 bp  
 \* 18595 19287: contig of 693 bp in length  
 \* 19288 19387: gap of 100 bp  
 \* 19388 20085: contig of 698 bp in length  
 \* 20086 20185: gap of 100 bp  
 \* 20186 20933: contig of 748 bp in length  
 \* 20934 21033: gap of 100 bp  
 \* 21034 21782: contig of 749 bp in length  
 \* 21783 21882: gap of 100 bp  
 \* 21883 22592: contig of 710 bp in length  
 \* 22593 22692: gap of 100 bp  
 \* 22693 23421: contig of 729 bp in length  
 \* 23422 23521: gap of 100 bp  
 \* 23522 24215: contig of 694 bp in length  
 \* 24216 24315: gap of 100 bp  
 \* 24316 25019: contig of 704 bp in length  
 \* 25020 25119: gap of 100 bp  
 \* 25120 25837: contig of 718 bp in length.

## FEATURES

## source

1. 25837  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP11-5803"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 693..752  
 /estimated\_length=100  
 1496..1595  
 /estimated\_length=100  
 2283..2382  
 /estimated\_length=100  
 3096..3195  
 /estimated\_length=100  
 3897..3996  
 /estimated\_length=100  
 4696..4795  
 /estimated\_length=100  
 5527..5626  
 /estimated\_length=100  
 6340..6439  
 /estimated\_length=100  
 7143..7242  
 /estimated\_length=100  
 7934..8033  
 /estimated\_length=100  
 8750..8849  
 /estimated\_length=100  
 9561..9660  
 /estimated\_length=100  
 10366..10465  
 /estimated\_length=100  
 11187..11286

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

## gap

gap /estimated\_length=100  
 gap 12000..12099  
 gap /estimated\_length=100  
 gap 12810..12909  
 gap /estimated\_length=100  
 gap 13609..13708  
 gap /estimated\_length=100  
 gap 14405..14504  
 gap /estimated\_length=100  
 gap 15225..15324  
 gap /estimated\_length=100  
 gap 16049..16148  
 gap /estimated\_length=100  
 gap 16865..16964  
 gap /estimated\_length=100  
 gap 17678..17777  
 gap /estimated\_length=100  
 gap 18495..18594  
 gap /estimated\_length=100  
 gap 19288..19387  
 gap /estimated\_length=100  
 gap 20086..20185  
 gap /estimated\_length=100  
 gap 20934..21033  
 gap /estimated\_length=100  
 gap 21783..21882  
 gap /estimated\_length=100  
 gap 22593..22692  
 gap /estimated\_length=100  
 gap 23422..23521  
 gap /estimated\_length=100  
 gap 24216..24315  
 gap /estimated\_length=100  
 gap 25020..25119  
 gap /estimated\_length=100

## ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 25837;  
 Best Local Similarity 88.0%; Pred. No. 95;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGCTGGAATCTCAGGCTGAGA 25  
 |||||  
 Db 2847 AGAAGCTGGAATCTCAGGCTGAAA 2823

RESULT 8  
 AL596124/c  
 LOCUS Human DNA sequence from clone RP11-141H1 on chromosome 10, complete sequence.  
 DEFINITION  
 ACCESSION AL596124.6 GI:15131373  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 30005)  
 AUTHORS Clark.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 COMMENT On Aug 9, 2001 this sequence version replaced gi:15041967.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-141H1 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk  
 -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES  
 source

1..30005  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-141H1"  
 /clone\_lib="RPCI-11.1"  
 2000  
 /note="Clone\_right\_end: RP11-347I122"  
 28006  
 /note="Clone\_left\_end: RP11-730A19"

## misc\_feature

## misc\_feature

## ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 30005;  
 Best Local Similarity 88.0%; Pred. No. 94;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGCTGGAATCTCAGGCTGAGA 25  
 |||||  
 Db 4252 AAATGACTGGAATCTCAGGCGAGAGA 4228

## RESULT 9

## AL354949/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## Direct Submission

## Submitted (13-MAY-2005)

## Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

## Clone requests: clonerequest@sanger.ac.uk

## On Oct 29, 2001 this sequence version replaced gi:14970325.

## The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

## Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at

## http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

## http://www.sanger.ac.uk/HGP/Chr1

## RP11-82L20 is from the library RPCI-11.1 constructed by the group

of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one subclone; and the assembly was confirmed by restriction digest,  
 except on the rare occasion of the clone being a YAC.

## FEATURES

```

source
  1. .180569
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="1"
    /clone_lib="RPC1-11.1"
    /notes="Clone left_end: RP11-82L20"
    /locus_tag="RP11-82L20"
    /product="neural growth regulator 1"
    /protein_id="CAI16869.1"
    /db_xref="GI:55958413"
    /db_xref="InterPro:IPR003598"
    /db_xref="InterPro:IPR003599"
    /db_xref="InterPro:IPR007110"
    /db_xref="UniProt/TREMBL:Q5VT21"
    /translation="MDMLLVGACSNQWLAALLSLCLLPSCLPAGSVDFPAA
VNMVVRKGDITAVLCYLEDGASGLNRSLIFAGDKWSVDPKVSISTLNKRDYS
LQIQNVDTDDGYTCSVQTHPTMQVHLTVQVPPKIYDISNDMTVAEGTNVLTVC
LATGKPEPSISWRHISPSAKPFENGQYLDIYIGTRDQAGEYECSAENDVSFPDVRKVK
VYVNFAPTIQEIKSGTVTPGRSLIRCEGAGVPPPAFVYKGEKLFNGQOQGIIONF
VRSILTVTNVQEHFGNYTCVAANKLGTNASLPLNPSTAGYGITGSADVLFSWCY
LVLTSSTSFYFLKNAILQ"
    join(complement(AL355660.13:129837..6022),
    complement(AL355660.13:129837..130069),
    complement(AL359821.11:86881..87006),
    complement(AL359821.11:8717..8848),
    complement(AL36081..136201), complement(117872..118023),
    complement(AL627317.7:19630..24258))
    /gene="NEGR1"
    /locus_tag="RP11-316C12.1-001"
    /standard_name="OTHUMP00000011267"
    /note="match: proteins: Tr:CAD97961 Tr:Q8N440 Tr:Q8NAQ3"
    /codon_start=1
    /product="neural growth regulator 1"
    /protein_id="CAI16869.1"
    /db_xref="GI:55958413"
    /db_xref="InterPro:IPR003598"
    /db_xref="InterPro:IPR003599"
    /db_xref="InterPro:IPR007110"
    /db_xref="UniProt/TREMBL:Q5VT21"
    /translation="MDMLLVGACSNQWLAALLSLCLLPSCLPAGSVDFPAA
VNMVVRKGDITAVLCYLEDGASGLNRSLIFAGDKWSVDPKVSISTLNKRDYS
LQIQNVDTDDGYTCSVQTHPTMQVHLTVQVPPKIYDISNDMTVAEGTNVLTVC
LATGKPEPSISWRHISPSAKPFENGQYLDIYIGTRDQAGEYECSAENDVSFPDVRKVK
VYVNFAPTIQEIKSGTVTPGRSLIRCEGAGVPPPAFVYKGEKLFNGQOQGIIONF
VRSILTVTNVQEHFGNYTCVAANKLGTNASLPLNPSTAGYGITGSADVLFSWCY
LVLTSSTSFYFLKNAILQ"
    join(complement(AL355660.13:129837..129861),
    complement(AL359821.11:86881..87006),
    complement(AL359821.11:8717..8848),
    complement(136081..136201), complement(117872..118023),
    complement(AL627317.7:24134..24258))
    /gene="NEGR1"
    /locus_tag="RP11-316C12.1-002"
    /standard_name="OTHUMP00000011268"
    /product="neural growth regulator 1"
    /protein_id="CAI16870.1"
    /db_xref="GI:55958414"
    /db_xref="InterPro:IPR003598"
    /db_xref="InterPro:IPR007110"
    /db_xref="UniProt/TREMBL:Q8N440"
    /translation="MQVHLTVQVPPKIYDISNDMTVAEGTNVLTCLATGKPEPSISM
RHISPSAKPFENGQYLDIYIGTRDQAGEYECSAENDVSFPDVRKVKVNFAPTIQEI
KSGTVTPGRSLIRCEGAGVPPPAFVYKGEKLFNGQOQGIIONFSTSLITVNTV
QGHFGNYTCVAANKLGTNASLPLNPSTAGYGITGSADVLFSWCYLVLTSSFTSIF
YLKNAILQ"
    misc_feature
      180569
    /note="Clone_right_end: RP11-82L20"
  ORIGIN
    Query Match      80.8%; Score 20.2; DB 8; Length 180569;
    Best Local Similarity 88.0%; Pred. No. 84;
    Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

  QY  1  AAAAACTGGAATCTCAGGCTGAGA 25
      |||||
  Db   14388 AGAAACTGGAATCTCAGGCTGAAA 14364

  RESULT 10
  AC156206/c
  LOCUS
  DEFINITION
    AC156206
    Bos taurus clone CH240-39N6, *** SEQUENCING IN PROGRESS ***, 10
    unordered pieces.
  ACCESSION
    AC156206
  VERSION
    AC156206.2
  KEYWORDS
    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
  SOURCE
    Bos taurus (cow)
  ORGANISM
    Bos taurus

```







```

/estimated_length=unknown
19244..192343
/estimated_length=unknown

ORIGIN
Query Match      80.8%; Score 20.2; DB 14; Length 193379;
Best Local Similarity 88.0%; Pred. No. 83;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGGATCTCAGGCTGAGA 25
|||||
DB 54135 AAAAACTGGACCTCAAACTGAGA 54111

RESULT 11
AC053464/c
LOCUS          194019 bp      DNA      linear      HTG 30-AUG-2001
DEFINITION    Homo sapiens chromosome 10 clone RP11-383B5, WORKING DRAFT
SEQUENCE      12 unordered pieces.
AC053464
AC053464.4 GI:13786255
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE        Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE     Smith,D.R.
AUTHORS       Genome Therapeutics Corporation Sequencing Center: Human Genome
TITLE         Sequence Data
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 194019)
AUTHORS       Smith,D.R.
TITLE         Direct Submission
JOURNAL       Submitted (16-APR-2000) Genome Therapeutics Corporation, 100 Beaver
COMMENT       Street, Waltham, MA 02453, USA
On Apr 25, 2001 this sequence version replaced gi:10944437.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg209
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 189621 bases at least Q40
Consensus quality: 191316 bases at least Q30
Consensus quality: 192144 bases at least Q20
Insert size: 196098; sum-of-contigs
Quality coverage: 6.1x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1409: contig of 1409 bp in length
* 1509: gap of unknown length
* 1510: contig of 1785 bp in length
* 3295: gap of unknown length
* 3395: contig of 2024 bp in length
* 5418: gap of unknown length
* 5519: contig of 1048 bp in length
* 6567: gap of unknown length
* 6667: contig of 3193 bp in length
* 9860: gap of unknown length

* 9960 15190: contig of 5231 bp in length
* 15191 15290: gap of unknown length
* 15291 25306: contig of 10016 bp in length
* 25307 25406: gap of unknown length
* 25407 39785: contig of 14379 bp in length
* 39786 39885: gap of unknown length
* 39886 57462: contig of 17577 bp in length
* 57463 57562: gap of unknown length
* 57563 79694: contig of 22131 bp in length
* 79694 79793: gap of unknown length
* 79794 122822: contig of 43029 bp in length
* 122823 122923: gap of unknown length
* 122923 194019: contig of 71097 bp in length.

FEATURES
Location/Qualifiers
1..194019
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPC1-11"
/clone="RP11-383B5"
1..1409
/notes="assembly_name:Contig7"
1410..1509
/estimated_length=unknown
1510..3294
/notes="assembly_name:Contig12"
3295..3394
/estimated_length=unknown
3395..5418
/notes="assembly_name:Contig13"
5419..5518
/estimated_length=unknown
5519..6567
/notes="assembly_name:Contig14"
6567..6666
/estimated_length=unknown
6667..9859
/notes="assembly_name:Contig16"
9860..9959
/estimated_length=unknown
9960..15190
/notes="assembly_name:Contig17"
15191..15290
/estimated_length=unknown
15291..25306
/notes="assembly_name:Contig18"
25307..25406
/estimated_length=unknown
25407..39785
/notes="assembly_name:Contig19"
39786..39885
/estimated_length=unknown
39886..57462
/notes="assembly_name:Contig20"
57463..57562
/estimated_length=unknown
57563..79693
/notes="assembly_name:Contig21"
79694..79793
/estimated_length=unknown
79794..122822
/notes="assembly_name:Contig22"
122823..122922
/estimated_length=unknown
122923..194019
/notes="assembly_name:Contig23"

ORIGIN
Query Match      80.8%; Score 20.2; DB 14; Length 194019;
Best Local Similarity 88.0%; Pred. No. 83;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
    ||| ||||| ||||| ||||| |||||
Db 94166 AATGACTGGAATCTCAGGCAGAGA 94142

RESULT 12
LOCUS AC013446
DEFINITION Homo sapiens chromosome 10 clone RP11-513P21, WORKING DRAFT
ACCESSION AC013446
VERSION AC013446.3 GI:7923997
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 208202)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208202)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 18, 2000 this sequence version replaced gi:6850545.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0513P21
----- Summary_Statistics -----
Sequencing vector: M13; 84%
Sequencing vector: plasmid; 16%
Chemistry: Dye-primer ET; 84% of reads
Chemistry: Dye-terminator Big Dye; 16% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 196339 bases at least Q40
Consensus quality: 199886 bases at least Q30
Consensus quality: 202038 bases at least Q20
Insert size: 215000; agarose-fp
Insert size: 206002; sum-of-contigs
Quality coverage: 4.38 in Q20 bases; agarose-fp
Quality coverage: 4.26 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1699: contig of 1699 bp in length
* 1700: gap of unknown length
* 1800: contig of 1441 bp in length
* 3241: gap of unknown length
* 3341: contig of 2875 bp in length
* 6216: gap of unknown length
* 6316: contig of 2740 bp in length
* 9056: gap of unknown length
* 9156: contig of 2689 bp in length
* 11845: gap of unknown length
* 11945: contig of 3647 bp in length
* 15591: gap of unknown length
* 15692: contig of 2566 bp in length
* 18258: gap of unknown length
* 18358: contig of 5081 bp in length
* 23438: gap of unknown length

23538: gap of unknown length
27553: contig of 4015 bp in length
27653: gap of unknown length
32488: contig of 4835 bp in length
32588: gap of unknown length
36405: contig of 3817 bp in length
36505: gap of unknown length
40593: contig of 4088 bp in length
40693: gap of unknown length
47782: contig of 7089 bp in length
47882: gap of unknown length
55184: contig of 7302 bp in length
55284: gap of unknown length
63288: contig of 8044 bp in length
63428: gap of unknown length
74101: contig of 10673 bp in length
74201: gap of unknown length
85050: contig of 10849 bp in length
85150: gap of unknown length
100569: contig of 15419 bp in length
100669: gap of unknown length
116236: contig of 15567 bp in length
116336: gap of unknown length
131648: contig of 15312 bp in length
131748: gap of unknown length
148926: contig of 17078 bp in length
148926: gap of unknown length
175149: contig of 28223 bp in length
175150: gap of unknown length
208202: contig of 32953 bp in length.

Location/Qualifiers
1..208202
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-513P21"

misc_feature 1..1699
/note="assembly_name:Contig10"
gap 1700..1799
/estimated_length=unknown
misc_feature 1800..3240
/note="assembly_name:Contig11"
gap 3241..3340
/estimated_length=unknown
misc_feature 3341..6215
/note="assembly_name:Contig12"
gap 6216..6315
/estimated_length=unknown
misc_feature 6316..9055
/note="assembly_name:Contig13"
gap 9056..9155
/estimated_length=unknown
misc_feature 9156..11844
/note="assembly_name:Contig14"
gap 11845..11944
/estimated_length=unknown
misc_feature 11945..15591
/note="assembly_name:Contig15"
gap 15592..15691
/estimated_length=unknown
misc_feature 15692..18257
/note="assembly_name:Contig16"
gap 18258..18357
/estimated_length=unknown
misc_feature 18358..23438
/note="assembly_name:Contig17"
gap 23439..23538
/estimated_length=unknown
misc_feature 23539..27553
/note="assembly_name:Contig18"
gap 27554..27653
/estimated_length=unknown
```

```

misc_feature 27654..32488
              /note="assembly_name:Contig19"
gap          32489..32588
              /estimated_length=unknown
misc_feature 32589..36405
              /note="assembly_name:Contig20"
gap          36406..36505
              /estimated_length=unknown
misc_feature 36506..40593
              /note="assembly_name:Contig21
              clone_end:T7
              vector_side:right"
gap          40594..40693
              /estimated_length=unknown
misc_feature 40694..47782
              /note="assembly_name:Contig22"
gap          47783..47882
              /estimated_length=unknown
misc_feature 47883..55184
              /note="assembly_name:Contig23"
gap          55185..55284
              /estimated_length=unknown
misc_feature 55285..63328
              /note="assembly_name:Contig24
              clone_end:SP6
              vector_side:left"
gap          63329..63428
              /estimated_length=unknown
misc_feature 63429..74101
              /note="assembly_name:Contig25"
gap          74102..74201
              /estimated_length=unknown
misc_feature 74202..85050
              /note="assembly_name:Contig26"
gap          85051..85150
              /estimated_length=unknown
misc_feature 85151..100569
              /note="assembly_name:Contig27"
gap          100570..100669
              /estimated_length=unknown
misc_feature 100670..116236
              /note="assembly_name:Contig28"
gap          116237..116336
              /estimated_length=unknown
misc_feature 116337..131648
              /note="assembly_name:Contig29"
gap          131649..131748
              /estimated_length=unknown
misc_feature 131749..148826
              /note="assembly_name:Contig30"
gap          148827..148926
              /estimated_length=unknown
misc_feature 148927..175149
              /note="assembly_name:Contig31"
gap          175150..175249
              /estimated_length=unknown
misc_feature 175250..208202
              /note="assembly_name:Contig32"

```

## ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 208202;  
 Best local Similarity 88.0%; Pred. No. 83;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAGCTGGAATCTCAGGCTGAGA 25

Db 142977 AAATGACTGGAATCTCAGGCTGAGA 143001

## RESULT 13

AC152756/c 231342 bp DNA linear HTG 01-JUL-2005  
 LOCUS  
 DEFINITION Bos taurus clone CH240-17N21, WORKING DRAFT SEQUENCE, 20 unordered

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

pieces.  
 AC152756.3 GI:68268033  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 Bos taurus (cow)  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Pecora; Bovidae; Bovinae; Bos.  
 1 (bases 1 to 231342)  
 Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Iolebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhea, L., Lousegod, H., Lozado, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
 Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,  
 Puar, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S. J.,  
 Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J.,  
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wlsczyk, R., Woodden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstock, G. and Gibbs, R. A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 231342)  
 Worley, K. C.  
 Direct Submission  
 Submitted (18-NOV-2004) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 231342)  
 Cow Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 28, 2005 this sequence version replaced gi:58037610.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: FAXI  
Center clone name: CH240-17N21  
----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 224136 bases at least Q40  
Consensus quality: 225748 bases at least Q30  
Consensus quality: 227236 bases at least Q20  
Estimated insert size: 226781; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1302: contig of 1302 bp in length  
\* 1303 2070: gap of 768 bp  
\* 2071 4765: contig of 2635 bp in length  
\* 4766 4815: gap of 50 bp  
\* 4816 22885: contig of 18070 bp in length  
\* 22886 22936: gap of 50 bp  
\* 22936 36075: contig of 13140 bp in length  
\* 36076 36125: gap of 50 bp  
\* 36126 69777: contig of 33652 bp in length  
\* 69778 69827: gap of 50 bp  
\* 69828 97935: contig of 28108 bp in length  
\* 97936 97986: gap of 50 bp  
\* 97986 101342: contig of 3357 bp in length  
\* 101343 101392: gap of 50 bp  
\* 101393 105292: contig of 3900 bp in length  
\* 105293 105342: gap of 50 bp  
\* 105343 137936: contig of 32594 bp in length  
\* 137937 137986: gap of 50 bp  
\* 137987 143769: contig of 5783 bp in length  
\* 143770 143819: gap of 50 bp  
\* 143820 145030: contig of 1211 bp in length  
\* 145031 145080: gap of 50 bp  
\* 145081 150012: contig of 4932 bp in length  
\* 150013 150062: gap of 50 bp  
\* 150063 154610: contig of 4548 bp in length  
\* 154611 154660: gap of 50 bp  
\* 154661 165400: contig of 10740 bp in length  
\* 165401 165450: gap of 50 bp  
\* 165451 179172: contig of 13722 bp in length  
\* 179173 179222: gap of 50 bp  
\* 179223 224914: contig of 45692 bp in length  
\* 224915 224964: gap of 50 bp  
\* 224965 226156: contig of 1192 bp in length  
\* 226157 226256: gap of unknown length

\* 226257 227672: contig of 1416 bp in length  
\* 227673 227772: gap of unknown length  
\* 227773 228777: contig of 1005 bp in length  
\* 228778 228877: gap of unknown length  
\* 228878 231342: contig of 2465 bp in length.

FEATURES  
source

1. .231342  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-17N21"  
1303. .2070  
/estimated\_length=768  
4766. .4815  
/estimated\_length=50  
22886. .22935  
/estimated\_length=50  
36076. .36125  
/estimated\_length=50  
69778. .69827  
/estimated\_length=50  
97936. .97985  
/estimated\_length=50  
101343. .101392  
/estimated\_length=50  
105293. .105342  
/estimated\_length=50  
137937. .137986  
/estimated\_length=50  
143770. .143819  
/estimated\_length=50  
145031. .145080  
/estimated\_length=50  
150013. .150062  
/estimated\_length=50  
154611. .154660  
/estimated\_length=50  
165401. .165450  
/estimated\_length=50  
179173. .179222  
/estimated\_length=50  
224915. .224964  
/estimated\_length=50  
226157. .226256  
/estimated\_length=unknown  
227673. .227772  
/estimated\_length=unknown  
228778. .228877  
/estimated\_length=unknown

ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 231342;  
Best Local Similarity 88.0%; Pred. No. 82;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||

Db 93857 AAAAACTGGAACCTCATGCTGAGA 93833  
|||||

RESULT 14

BD212075

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 300)

BD212075 Novel human genes and gene expression products ii.  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

BD212075.1 GI:33021845  
JP 2002519000-A/217.

Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

1 (bases 1 to 300)

AUTHORS Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J.S., Reinhard, C., Giese, K., Randazzo, F., Kennedy, G.C., Pot, D., Kassam, A., Lamson, G., Drmanac, R., Crkvenjakov, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D., Kita, D., Garcia, V., Jones, L.W. and Crain, B.S.

TITLE Novel human genes and gene expression products ii

JOURNAL Patent: JP 2002519000-A 217 02-JUL-2002;

COMMENT CHIRON CORP. HYSEQ INC

OS Homo sapiens (human)

PN JP 2002519000-A/217

PD 02-JUL-2002

PF 28-JAN-1999 JP 2000556580

PR 28-JAN-1998 US 60/072910, 24-FEB-1998 US 60/075954 PR

31-MAR-1998 US 60/080114, 03-APR-1998 US 60/080515 PR

03-APR-1998 US 60/080666, 21-OCT-1998 US 60/105234 PR

28-OCT-1998 US 60/105877

PI LOUIS T WILLIAMS, JAIME ESCOBEDO, MICHAEL A INNIS, PABLO PI

DOMINGUEZ GARCIA,

PI JULIE SUDDUTH KLINGER, CHRISTOPH REINHARD, KLAUSE GIESE, FILIPPO

PI RANDAZZO,

PI GIULIA C KENNEDY, DAVID POT, ALTAF KASSAM, GEORGE LAMSON, RADOJE

PI DRMANAC,

PI RADOMIR CRKVENJAKOV, MARK DICKSON, SNEZANA DRMANAC, IVAN LABAT,

PI DENA LESHKOWITZ, DAVID KITA, VERONICA GARCIA, LEE WILLIAM JONES,

PI BIRJIT STACHE CRAIN

PC C12N15/09, C12N15/09, C07K14/47, C07K14/82, C07K16/18, C12N1/15, PC

C12N1/19,

PC C12N1/21, C12N5/10, C12Q1/68, C12N15/00, C12N5/00, C12N15/00 CC

Novel human genes and gene expression products ii

FT source

FT Location/Qualifiers

1. .300

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 300;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGA 23

|||||

Db 104 AAACCTGGAATCTCAGGCTGA 123

RESULT 15

BD215003

LOCUS Novel human genes and gene expression products ii. 758 bp DNA linear PAT 17-JUL-2003

DEFINITION

ACCESSION BD215003

VERSION BD215003.1 GI:33024773

KEYWORDS JP 2002519000-A/3145.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 758)

AUTHORS Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J.S., Reinhard, C., Giese, K., Randazzo, F., Kennedy, G.C., Pot, D., Kassam, A., Lamson, G., Drmanac, R., Crkvenjakov, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D., Kita, D., Garcia, V., Jones, L.W. and Crain, B.S.

TITLE Novel human genes and gene expression products ii

JOURNAL Patent: JP 2002519000-A 3145 02-JUL-2002;

COMMENT CHIRON CORP. HYSEQ INC

OS Homo sapiens (human)

PN JP 2002519000-A/3145

PD 02-JUL-2002

PF 28-JAN-1999 JP 2000556580

PR 28-JAN-1998 US 60/072910, 24-FEB-1998 US 60/075954 PR

31-MAR-1998 US 60/080114, 03-APR-1998 US 60/080515 PR

03-APR-1998 US 60/080666, 21-OCT-1998 US 60/105234 PR

28-OCT-1998 US 60/105877

PI LOUIS T WILLIAMS, JAIME ESCOBEDO, MICHAEL A INNIS, PABLO PI

DOMINGUEZ GARCIA,

PI JULIE SUDDUTH KLINGER, CHRISTOPH REINHARD, KLAUSE GIESE, FILIPPO

PI RANDAZZO,

PI GIULIA C KENNEDY, DAVID POT, ALTAF KASSAM, GEORGE LAMSON, RADOJE

PI DRMANAC,

PI RADOMIR CRKVENJAKOV, MARK DICKSON, SNEZANA DRMANAC, IVAN LABAT,

PI DENA LESHKOWITZ, DAVID KITA, VERONICA GARCIA, LEE WILLIAM JONES,

PI BIRJIT STACHE CRAIN

PC C12N15/09, C12N15/09, C07K14/47, C07K14/82, C07K16/18, C12N1/15, PC

C12N1/19,

PC C12N1/21, C12N5/10, C12Q1/68, C12N15/00, C12N5/00, C12N15/00 CC

= A, T, C or G

FT Key Location/Qualifiers

FT misc feature (1). .(758).

1. .758

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 758;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGA 23

|||||

Db 148 AAACCTGGAATCTCAGGCTGA 167

RESULT 16

ABI68750

LOCUS Macaca fascicularis testis cDNA, clone: QtsA-14569, similar to human tubby like protein 4 (TULP4), mRNA, RefSeq: NM\_020245.2. 2272 bp mRNA linear PRI 18-JUN-2005

DEFINITION

ACCESSION ABI68750.1 GI:67969004

VERSION ABI68750.1

KEYWORDS Macaca fascicularis (crab-eating macaque)

SOURCE Macaca fascicularis

ORGANISM Macaca fascicularis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE

1 International consortium for macaque cDNA sequencing and analysis. DNA sequences of macaque genes expressed in brain or testis and its evolutionary implications

2 Unpublished

JOURNAL

REFERENCE

AUTHORS Oeada, N., Hirata, M., Tanuma, R., Kusuda, J., Hida, M., Suzuki, Y., Sugano, S., Gojobori, T., Shen, J.C.-K., Wu, C.I. and Hashimoto, K.

TITLE Substitution rate and structural divergence of 5'UTR evolution: Comparative analysis between human and cynomolgus monkey cDNAs

JOURNAL Unpublished

REFERENCE

3 (bases 1 to 2272)

AUTHORS Hashimoto, K., Kusuda, J. and Sugano, S.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-2004) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo, 162-8640, Japan

1-chome, Shinjuku-ku, Tokyo, 162-8640, Japan

(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)

The International consortium for macaque cDNA sequencing and analysis consists of: Department of Virology and Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; Division of Genetic Resources, National Institute of Infectious Diseases of Japan, Tokyo, Japan; National Health Research Institute, Taipei, Taiwan; Institute of Molecular Biology,



CDS

/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
1358..3394  
/notes="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAD34914.1"  
/db\_xref="GI:21439054"  
/translation="MYAAVEHGVLCSDSNILGWSKGRVPKSEKEKPCVRRYYEEG  
WLATNGRGVGVGTFTSSHCRDRSTPQRINFLRGHNSVVLVRWNEPYQKLATCDA  
DGGIFVWIOYEGRWSELVNDRGAQVSDFTWSHDGTQALLSYRDFGLVGSVSGORHW  
YPIFLVEDSSESDTSDDYAPPQDGAAYPIPVQNIKPLLTVSTSGDISLNNYDDL  
SPTVIRSGLKEVVAQWCTQGDLLAVAGMERQTLGELPGLKSAWVKFYNVRGEHI  
FLDTLVQRPILISICWHRDRLMASGPAVVRVHRVSSLQLLCOQAIASITLRED  
KDVSKLTLPRLCSYLSAFITPIKPIPDNNMRDFVSPSAGNERLHCTWKRTEDD  
PEVGGPCYTLLEYLGGVLVPLIKGRIRISKLRFVIMDRPTDSKDEIYGNLSISTVI  
DSCNCSDDSIELSDDAAKSPKISRASKSPKPRIISIEARKSPKPRAAQELSRSP  
RLPKRPSVGSPLSTRFPEDITQHNLAQVTSNIWGTNFKIYGLAAFLPTNLGAV  
LYKTSLLHLQPRQMTIYLPVVRKISMDYINLVPFNPNVSEDEDDLPGDVWTD"

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 5174;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 19

AF288480 5183 bp mRNA linear PRI 20-AUG-2000  
LOCUS Homo sapiens tubby super-family protein (TUSP) mRNA, complete cds,  
DEFINITION alternatively spliced.  
ACCESSION AF288480  
VERSION AF288480.1 GI:9858153  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 5183)  
Li, Q.-Z., Wang, C. and She, J.-X.  
Molecular cloning of TUSP gene, a novel member of tubby gene  
super-family, its expression, alternative splicing and chromosome  
localization on 6q25-26, from human and mouse  
Unpublished  
2 (bases 1 to 5183)  
Li, Q.-Z., Wang, C., Shi, J.-D., Cruze, P. and She, J.-X.  
Direct Submission  
Submitted (19-JUL-2000) Pathology, Immunology and Laboratory  
Science, University of Florida, 1600 SW Archer Road, Rm. D6-15,  
Gainesville, FL 32610, USA  
FEATURES  
source  
1..5183  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6q25-q26"  
1..5183  
/gene="TUSP"  
1358..3394  
/gene="TUSP"  
/note="alternatively spliced"  
/codon\_start=1  
/product="tubby super-family protein"  
/protein\_id="AA01020.1"  
/db\_xref="GI:9858154"  
/translation="MYAAVEHGVLCSDSNILGWSKGRVPKSEKEKPCVRRYYEEG  
WLATNGRGVGVGTFTSSHCRDRSTPQRINFLRGHNSVVLVRWNEPYQKLATCDA

ORIGIN

Query Match 80.0%; Score 20; DB 8; Length 5183;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 20

CS033671 11127 bp DNA linear PAT 10-MAR-2005  
LOCUS Sequence 3177 from Patent WO2005016962.  
DEFINITION  
ACCESSION CS033671  
VERSION CS033671.1 GI:60732618  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
Abbas, A., Clark, H., Ouyang, W., Williams, M. P., Wood, W. I. and Wu, T. D.  
Compositions and methods for the treatment of immune related  
diseases  
Patent: WO 2005016962-A 3177 24-FEB-2005;  
Genentech, Inc. (US)  
FEATURES  
source  
1..11127  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 11127;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 21

CS042623 11127 bp DNA linear PAT 22-MAR-2005  
LOCUS Sequence 3177 from Patent WO2005019258.  
DEFINITION  
ACCESSION CS042623  
VERSION CS042623.1 GI:61849791  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
Abbas, A., Clark, H., Ouyang, W., Williams, P. M., Wood, W. I. and Wu, T. D.  
Compositions and methods for the treatment of immune related  
diseases  
Patent: WO 2005019258-A 3177 03-MAR-2005;  
Genentech, Inc. (US)  
JOURNAL

CDS

/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
1358..3394  
/notes="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAD34914.1"  
/db\_xref="GI:21439054"  
/translation="MYAAVEHGVLCSDSNILGWSKGRVPKSEKEKPCVRRYYEEG  
WLATNGRGVGVGTFTSSHCRDRSTPQRINFLRGHNSVVLVRWNEPYQKLATCDA  
DGGIFVWIOYEGRWSELVNDRGAQVSDFTWSHDGTQALLSYRDFGLVGSVSGORHW  
YPIFLVEDSSESDTSDDYAPPQDGAAYPIPVQNIKPLLTVSTSGDISLNNYDDL  
SPTVIRSGLKEVVAQWCTQGDLLAVAGMERQTLGELPGLKSAWVKFYNVRGEHI  
FLDTLVQRPILISICWHRDRLMASGPAVVRVHRVSSLQLLCOQAIASITLRED  
KDVSKLTLPRLCSYLSAFITPIKPIPDNNMRDFVSPSAGNERLHCTWKRTEDD  
PEVGGPCYTLLEYLGGVLVPLIKGRIRISKLRFVIMDRPTDSKDEIYGNLSISTVI  
DSCNCSDDSIELSDDAAKSPKISRASKSPKPRIISIEARKSPKPRAAQELSRSP  
RLPKRPSVGSPLSTRFPEDITQHNLAQVTSNIWGTNFKIYGLAAFLPTNLGAV  
LYKTSLLHLQPRQMTIYLPVVRKISMDYINLVPFNPNVSEDEDDLPGDVWTD"

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 5174;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 19

AF288480 5183 bp mRNA linear PRI 20-AUG-2000  
LOCUS Homo sapiens tubby super-family protein (TUSP) mRNA, complete cds,  
DEFINITION alternatively spliced.  
ACCESSION AF288480  
VERSION AF288480.1 GI:9858153  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 5183)  
Li, Q.-Z., Wang, C. and She, J.-X.  
Molecular cloning of TUSP gene, a novel member of tubby gene  
super-family, its expression, alternative splicing and chromosome  
localization on 6q25-26, from human and mouse  
Unpublished  
2 (bases 1 to 5183)  
Li, Q.-Z., Wang, C., Shi, J.-D., Cruze, P. and She, J.-X.  
Direct Submission  
Submitted (19-JUL-2000) Pathology, Immunology and Laboratory  
Science, University of Florida, 1600 SW Archer Road, Rm. D6-15,  
Gainesville, FL 32610, USA  
FEATURES  
source  
1..5183  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6q25-q26"  
1..5183  
/gene="TUSP"  
1358..3394  
/gene="TUSP"  
/note="alternatively spliced"  
/codon\_start=1  
/product="tubby super-family protein"  
/protein\_id="AA01020.1"  
/db\_xref="GI:9858154"  
/translation="MYAAVEHGVLCSDSNILGWSKGRVPKSEKEKPCVRRYYEEG  
WLATNGRGVGVGTFTSSHCRDRSTPQRINFLRGHNSVVLVRWNEPYQKLATCDA

ORIGIN

Query Match 80.0%; Score 20; DB 8; Length 5183;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 20

CS033671 11127 bp DNA linear PAT 10-MAR-2005  
LOCUS Sequence 3177 from Patent WO2005016962.  
DEFINITION  
ACCESSION CS033671  
VERSION CS033671.1 GI:60732618  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
Abbas, A., Clark, H., Ouyang, W., Williams, M. P., Wood, W. I. and Wu, T. D.  
Compositions and methods for the treatment of immune related  
diseases  
Patent: WO 2005016962-A 3177 24-FEB-2005;  
Genentech, Inc. (US)  
FEATURES  
source  
1..11127  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 80.0%; Score 20; DB 6; Length 11127;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTGGATCTCAGGCTGA 23  
|||||  
Db 964 AACTGGATCTCAGGCTGA 983

RESULT 21

CS042623 11127 bp DNA linear PAT 22-MAR-2005  
LOCUS Sequence 3177 from Patent WO2005019258.  
DEFINITION  
ACCESSION CS042623  
VERSION CS042623.1 GI:61849791  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
Abbas, A., Clark, H., Ouyang, W., Williams, P. M., Wood, W. I. and Wu, T. D.  
Compositions and methods for the treatment of immune related  
diseases  
Patent: WO 2005019258-A 3177 03-MAR-2005;  
Genentech, Inc. (US)  
JOURNAL



FEATURES	Location/Qualifiers		Query Match		80.0%; Score 20; DB 6; Length 11127;	
	source	1..11127	Best Local Similarity	100.0%; Pred. No. 1.2e+02;	Mismatches	0; Indels 0; Gaps 0;
ORIGIN	/organism="Homo sapiens"		Matches 20; Conservative 0;		Mismatches 0; Indels 0; Gaps 0;	
	/mol_type="unassigned DNA"		Db	964	AAACTGGAATCTCAGGCTGA 983	
ORIGIN	/db_xref="taxon:9606"		Qy		4 AAACCTGGAATCTCAGGCTGA 23	
			Db	964	AAACTGGAATCTCAGGCTGA 983	
ORIGIN			RESULT 22		AL360169	
			LOCUS	AF219946	11127 bp mRNA linear PRI 16-OCT-2001	
ORIGIN			DEFINITION		Homo sapiens tubby super-family protein (TUSP) mRNA, complete cds.	
			ACCESSION	AF219946		
ORIGIN			VERSION		AF219946.2 GI:12232630	
			KEYWORDS	Homo sapiens		
ORIGIN			SOURCE		Homo sapiens (human)	
			ORGANISM	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.	
ORIGIN			REFERENCE		1 (bases 1 to 11127)	
			AUTHORS	Li, Q., Wang, C., Ruan, Q., Shi, J., Yang, P., Eckenrode, S., Davoodi-Semirami, A., Kukar, T., Gu, Y., Lian, W., Wu, D. and She, J. X.	Molecular cloning and characterization of the mouse and human TUSP gene, a novel member of the tubby superfamily	
ORIGIN			JOURNAL		Gene 273 (2), 275-284 (2001)	
			PUBMED	11595174		
ORIGIN			REFERENCE		2 (bases 1 to 11127)	
			AUTHORS	Li, Q., Wang, C., Ruan, Q., Shi, J., Yang, P., Eckenrode, S. and She, J.	Direct Submission	
ORIGIN			TITLE		Submitted (28-DEC-1999) Department of Pathology, University of Florida, 1600 SW Archer Rd., Rm. D6-15, Gainesville, FL 32610, USA	
			JOURNAL	3 (bases 1 to 11127)		
ORIGIN			REFERENCE		Li, Q., Wang, C., Ruan, Q., Shi, J., Yang, P., Eckenrode, S. and She, J.	
			AUTHORS	Direct Submission		
ORIGIN			TITLE		Submitted (16-JAN-2001) Department of Pathology, University of Florida, 1600 SW Archer Rd., Rm. D6-15, Gainesville, FL 32610, USA	
			JOURNAL	Sequence update by submitter		
ORIGIN			REMARK		On Jan 16, 2001 this sequence version replaced gi:9502081.	
			COMMENT	Location/Qualifiers		
ORIGIN			FEATURES		1..11127	
			source	/organism="Homo sapiens"	/mol_type="mRNA"	
ORIGIN			gene			/db_xref="taxon:9606"
				/chromosomes="6"		
ORIGIN			CDS		/map="1.01 cr from NIB330, 2.63 cr from WI-4986"	
					/tissue type="brain"	
ORIGIN					1..11127	
					/gene="TUSP"	
ORIGIN					1358..5992	
					/genes="TUSP"	
ORIGIN					/codon start=1	
					/product="tubby super-family protein"	
ORIGIN					/protein_id="AAF87975.1"	
					/db_xref="GI:9502082"	
ORIGIN					/translation="MYAAVEHGFVLCSDSNILCLSKWGVKSEKPKVRRRYVEEG WLATNGRQVGVGTSTSHCRDRSTPQINFNLGRHNSVVLVRWNEPYQKLATCDA DGIIFWIOVEGRWSVELVNDRAQVSDFTWSDGTQALISVRDGFVLGVSGQORHW SEINLESITGIGTWPDDQVLFCTADQGVIMDCHGMLAHVLLHSDGVLGWSN YPILFVDSSESDDDDTAPFDGPAATPIQNIKPLLTVSTFTSGDISLNMVDDL SPTVRISGLKVVAQMCQTQDLLAVAGMERQTLGELPNGLLKSAWKFVNRGEHI FLDITLIVORPISICWGRDRSLMAGSPALYVVRVHRVSSLIQLCCQAIATSLRED KQVSKTLPLRLCSYLSATPIPTKPIPDNNMRDVFYSPAGNERLHCTMKRTEDD PVGGPCVTLXLEYLGLVPLKGRIRSKLRPEFVIMDPTSDKPEIYGNLSITVI DSCNCSDDSDILSDDDAAKSPKISRASKSPKLPRISTEARKSPKLPRAAQELSRSP	



/note="Sequence from overlapping clone AL161721. Assembly confirmed by restriction digest data."

misc\_feature

75390..75577      80.0%;      Score 20;      DB 8;      Length 180635;  
confirmed by restriction digest data."

gene

Join(78483..78534,79069..79159)  
/locus tag="RP11-732M18.2-001"

mRNA

Join(78483..78534,79069..79159)  
/locus tag="RP11-732M18.2-001"

misc\_feature

92518  
/note="match: ESTs: B1463512"

Query Match

Best Local Similarity      100.0%;      Pred. No. 1e+02;      Length 180635;  
Matches      20;      Conservative      0;      Mismatches      0;      Indels      0;      Gaps      0;

Oy      4      AAATGGGAATCTCAGGCTGA      23

Db      163421      AAATGGGAATCTCAGGCTGA      163402

RESULT 24

AC108611

LOCUS

DEFINITION      AC108611      295920 bp      DNA      linear      HTG 08-OCT-2002  
Rattus norvegicus clone CH230-292P23, \*\*\* SEQUENCING IN PROGRESS

ACCESSION

AC108611.4      GI:23195549

VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.

1      (bases 1 to 295920)

REFERENCE

Alden, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebreyeorgis, E., Geer, K., Gill, R., Grady, M., Guerrero, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kowar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Loulseg, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Fu, L.-L.,  
Pua, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

```

* 274665 276601: contig of 1937 bp in length
* 276602 276701: gap of unknown length
* 288825 288825: contig of 12124 bp in length
* 288826 288925: gap of unknown length
* 288926 295920: contig of 6995 bp in length.
FEATURES
source
1. .295920
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-292P23"
1. .1408
/notes="wgs contig"
misc_feature
52118..52217
/estimated_length=unknown
gap
270694..270793
/estimated_length=unknown
gap
272144..272243
/estimated_length=unknown
gap
274565..274664
/estimated_length=unknown
gap
276602..276701
/estimated_length=unknown
gap
288826..288925
/estimated_length=unknown
ORIGIN
Query Match 80.0%; Score 20; DB 14; Length 295920;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGC 20
|||||
Db 120957 AAAAACTGGAATCTCAGGC 120976

RESULT 25
AC025192 134514 bp DNA linear HTG 27-MAR-2003
LOCUS Homo sapiens chromosome 8 clone RP11-509E2 map 8, 3 unordered
DEFINITION pieces.
AC025192
AC025192.4 GI:18875263
VERSION HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 134514)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-509E2
Unpublished
2 (bases 1 to 134514)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castie,A., Choepel,Y., Colangelo,W., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Howland,J., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

```

```

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 134514)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Meneus,L.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2002 this sequence version replaced gi:12313839.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5592
Center clone name: 509_E_2
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 17787: contig of 17787 bp in length
* 17788 17887: gap of 100 bp
* 17888 97189: contig of 79301 bp in length
* 97189 97288: gap of 100 bp
* 97289 134514: contig of 37236 bp in length.
FEATURES
source
1. .134514
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-509E2"
/clone_lib="RP11-509E2"
17788..17887
/estimated_length=100
97189..97288
/estimated_length=100
ORIGIN
Query Match 79.2%; Score 19.8; DB 14; Length 134514;

```

Best Local Similarity 91.3%; Pred. No. 1.3e+02; Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGAG 24  
|||||  
Db 128346 AAAGACTGGAACCTCAGGCTGAG 128368  
|||||

RESULT 26  
AC069005  
LOCUS  
DEFINITION Homo sapiens chromosome 8 clone RP11-712115, WORKING DRAFT  
AC069005  
ACCESSION AC069005.3 GI:8844180  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1 (bases 1 to 162560)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 162560)  
Waterston,R.H.  
Direct Submission  
Submitted (16-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jun 30, 2000 this sequence version replaced gi:8469066.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H NH0712115  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Chemistry: Dye-terminator; plasmid; 0%  
Chemistry: Dye-terminator; Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 145600 bases at least Q40  
Consensus quality: 151065 bases at least Q30  
Consensus quality: 153390 bases at least Q20  
Insert size: 173000; agarose-fp  
Insert size: 159360; sum-of-contigs  
Quality coverage: 3.35 in Q20 bases; agarose-fp  
Quality coverage: 3.73 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1484: contig of 1483 bp in length  
1584: gap of unknown length  
3194: contig of 1611 bp in length  
3294: gap of unknown length  
3395: contig of 1178 bp in length  
4472: gap of unknown length  
4572: gap of unknown length  
4733: contig of 1548 bp in length  
6121: gap of unknown length  
6221: contig of 1826 bp in length  
8047: gap of unknown length  
8147: contig of 1755 bp in length  
9902: 10001: gap of unknown length

10002 12500: contig of 2499 bp in length  
12501 12600: gap of unknown length  
12601 12666: contig of 2666 bp in length  
15267 15366: gap of unknown length  
15367 17549: contig of 2183 bp in length  
17550 20378: contig of 2729 bp in length  
20379 20478: gap of unknown length  
20479 23345: contig of 2867 bp in length  
23346 27122: contig of 3677 bp in length  
27123 30048: contig of 2826 bp in length  
30049 30148: gap of unknown length  
30149 32221: contig of 2072 bp in length  
32221 35967: contig of 3647 bp in length  
35968 36068: gap of unknown length  
36069 39730: contig of 3663 bp in length  
39731 43447: contig of 3617 bp in length  
43448 43547: gap of unknown length  
43548 47467: contig of 3920 bp in length  
47468 51880: contig of 4313 bp in length  
51881 57315: contig of 5335 bp in length  
57316 61833: contig of 4418 bp in length  
61834 65951: contig of 4018 bp in length  
65952 70869: contig of 4818 bp in length  
70870 70969: gap of unknown length  
70970 77012: contig of 6043 bp in length  
77013 82785: contig of 5673 bp in length  
82786 88405: contig of 5520 bp in length  
88406 93767: contig of 5262 bp in length  
93768 93868: contig of 5705 bp in length  
93869 99572: gap of unknown length  
99573 109490: contig of 9818 bp in length  
109491 119557: contig of 9967 bp in length  
119558 131547: contig of 11890 bp in length  
131548 148258: contig of 16611 bp in length  
148259 162560: contig of 14202 bp in length.  
148359 Location/Qualifiers  
1. .162560  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/clone="RP11-712115"  
1484. .1583  
/estimated\_length=unknown  
3195. .3294  
/estimated\_length=unknown  
4473. .4572  
/estimated\_length=unknown  
6121. .6220  
/estimated\_length=unknown  
8047. .8147  
/estimated\_length=unknown  
9902. .10001  
/estimated\_length=unknown  
12501. .12500

FEATURES  
source

gap

gap

gap

gap

gap

gap

gap





JOURNAL Submitted (07-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Jan 7, 2003 this sequence version replaced gi:27476180.  
 All repeats were identified using RepeatMasker:  
 Smith, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L21648  
 Center clone name: 421\_P\_23  
 -----

## FEATURES

source

## Location/Qualifiers

1. .167878  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosomes="8"  
 /map="8"  
 /clone="RP11-421P23"  
 /clone\_lib="RP11-421P23 Human Male BAC"  
 complement(31..462)  
 /rpt\_family="MUT1C"  
 complement(894..996)  
 /rpt\_family="L1ME2"  
 1036..1346  
 /rpt\_family="AluSx"  
 complement(3464..3610)  
 /rpt\_family="L3"  
 3612..3691  
 /rpt\_family="MIR91A"  
 complement(3889..4086)  
 /rpt\_family="L3"  
 complement(4429..4496)  
 /rpt\_family="MIR"  
 complement(4945..6217)  
 /rpt\_family="HSMAR2"  
 complement(6639..6936)  
 /rpt\_family="AluSx"  
 7496..7806  
 /rpt\_family="AluJb"  
 complement(8574..8660)  
 /rpt\_family="MIR3"  
 complement(9589..9697)  
 /rpt\_family="MIR"  
 9913..10223  
 /rpt\_family="AluY"  
 11320..11369  
 /rpt\_family="GAAA"  
 complement(11413..11593)  
 /rpt\_family="L1PA4"  
 11803..11885  
 /rpt\_family="TAGA"  
 12730..12830  
 /rpt\_family="L2"  
 13706..14005  
 /rpt\_family="AluSx"  
 15664..15770  
 /rpt\_family="MIR3"  
 16008..16046  
 /rpt\_family="L3"  
 16183..16300  
 /rpt\_family="MIR"  
 complement(19228..19381)  
 /rpt\_family="AluSg/x"  
 19858..21101  
 /rpt\_family="L1ME1"  
 21178..21327  
 /rpt\_family="L1ME1"  
 complement(21343..21852)

repeat\_region /rpt\_family="L1P4"  
 21853..22231  
 /rpt\_family="L1PAL0"  
 22235..22332  
 /rpt\_family="L1ME1"  
 complement(22344..22473)  
 /rpt\_family="FLAM\_C"  
 22508..22555  
 /rpt\_family="TA)n"  
 22561..22636  
 /rpt\_family="L1ME2"  
 22704..22828  
 /rpt\_family="L2"  
 23023..23203  
 /rpt\_family="L2"  
 23455..23508  
 /rpt\_family="TA)n"  
 23693..24079  
 /rpt\_family="MLT11"  
 24286..24330  
 /rpt\_family="GA-rich"  
 24427..24468  
 /rpt\_family="AT-rich"  
 24986..25013  
 /rpt\_family="(TG)n"  
 complement(25041..25110)  
 /rpt\_family="MIR"  
 25148..25191  
 /rpt\_family="polypurine"  
 25198..25254  
 /rpt\_family="CT-rich"  
 complement(26391..26594)  
 /rpt\_family="MLT12"  
 26702..27267  
 /rpt\_family="MLT1E"  
 27434..27468  
 /rpt\_family="(TA)n"  
 28895..29360  
 /rpt\_family="LTR33"  
 29723..35996  
 /rpt\_family="L1PA7"  
 complement(36434..36707)  
 /rpt\_family="AluJo"  
 37116..37136  
 /rpt\_family="AT-rich"

Query Match 79.2%; Score 19.8; DB 8; Length 167878;  
 Best Local Similarity 91.3%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGGAACTCTCAGGCTGAG 24  
 Db 37448 AGAATCTGGAATCTCAGGCTGAG 37470

## RESULT 28

AC015468 189662 bp DNA linear PRI 01-MAY-2001  
 LOCUS Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.  
 DEFINITION AC015468  
 ACCESSION AC015468  
 VERSION AC015468.5 GI:13899433  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 189662)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 8, clone RP11-369E15  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 189662)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,



Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 189662)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 1, 2001 this sequence version replaced gi:12313808.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

# TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L2466

Center clone name: 369\_E15

## FEATURES source

Location/Qualifiers

1. .189662  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-369E15"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .100  
/rpt\_family="MLT1A1"  
1594. .1741  
/rpt\_family="GA-rich"  
complement(2782. .2957)  
/rpt\_family="MIR"  
3311. .3471  
/rpt\_family="MIR"  
3964. .4067  
/rpt\_family="MIR"

repeat\_region  
complement(4528. .4771)  
/rpt\_family="L2"  
4776. .4871  
/rpt\_family="MER4A"  
4869. .5223  
/rpt\_family="MER4A"  
complement(5255. .5765)  
/rpt\_family="L2"  
complement(5805. .6003)  
/rpt\_family="THE1A"  
6004. .6035  
/rpt\_family="TTA)n"  
complement(6151. .6319)  
/rpt\_family="SVA"  
complement(6320. .6468)  
/rpt\_family="THE1A"  
8722. .8751  
/rpt\_family="TTTAA)n"  
9155. .9553  
/rpt\_family="MSTB"  
9599. .9982  
/rpt\_family="THE1C"  
complement(10113. .10320)  
/rpt\_family="AluY"  
10822. .10970  
/rpt\_family="MIR3"  
11210. .11215  
/note="<30 qual SNGL region"  
11794. .12080  
/rpt\_family="AluSx"  
12974. .13010  
/rpt\_family="(TG)n"  
complement(13091. .13304)  
/rpt\_family="MER106A"  
13481. .13858  
/rpt\_family="MLT1C"  
14025. .14260  
/rpt\_family="L2"  
14584. .14679  
/rpt\_family="MIR"  
14723. .14825  
/rpt\_family="L2"  
15167. .15332  
/rpt\_family="MER33"  
15436. .15788  
/rpt\_family="L2"  
complement(16078. .16165)  
/rpt\_family="AluJ/FRAM"  
16532. .17094  
/rpt\_family="MLT1F1"  
17653. .17674  
/rpt\_family="(TAA)n"  
complement(17675. .17892)  
/rpt\_family="MER7C"  
complement(17897. .18650)  
/rpt\_family="LTR17"  
complement(18651. .18996)  
/rpt\_family="MER7C"  
18999. .19103  
/rpt\_family="(TAA)n"  
complement(19203. .19283)  
/rpt\_family="MIR"  
20464. .20523  
/rpt\_family="A-rich"  
complement(21279. .21902)  
/rpt\_family="MER41B"  
21922. .22092  
/rpt\_family="MIR"  
22249. .22373  
/rpt\_family="MIR"  
23956. .24058  
/rpt\_family="L2"  
complement(24069. .24198)

```

repeat_region /rpt_family="MIR"
25149 .25213
/rpt_family="L2"
repeat_region /rpt_family="L2"
complement(25783). .26130)
repeat_region /rpt_family="MUT11"
complement(26174). .26539)
repeat_region /rpt_family="MUT1A1"
complement(28187). .28528)
repeat_region /rpt_family="AluY"
complement(28792). .29263)
repeat_region /rpt_family="Charlie8"
complement(30652). .30730)
repeat_region /rpt_family="MER53"
complement(30986). .31117)
repeat_region /rpt_family="LIME1"
31113. .31556
/rpt_family="LIME2"
31590. .32174
repeat_region /rpt_family="LIME2"
32162. .32588
/rpt_family="LIME3"
33242. .33490
repeat_region /rpt_family="MIR"
complement(34416). .34803)
repeat_region /rpt_family="MSTB"
complement(35044). .35350)
repeat_region /rpt_family="AluSg"
36407. .36695
repeat_region /rpt_family="AluSx"
complement(37670). .37790)
repeat_region /rpt_family="FLAM.A"
complement(38065). .38132)
repeat_region /rpt_family="L2"
complement(38504). .38590)
repeat_region /rpt_family="MIR"
39044. .39228
repeat_region /rpt_family="MIR3"
complement(40317). .40473)

```

Query Match 79.2%; Score 19.8; DB 8; Length 189662;

Best Local Similarity 91.3%; Pred. No. 1.3e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGAG 24

Db 165165 AGAATCTGAATCTCAGGCTGAG 165187

```

RESULT 29
AC087221/c
LOCUS AC087221 203690 bp DNA linear HTG 24-MAY-2002
DEFINITION Homo sapiens chromosome 8 clone RP11-712115 map 8, WORKING DRAFT
SEQUENCE, 34 ordered pieces.
ACCESSION AC087221
VERSION AC087221.2 GI:211166223
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 203690)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,S.,
TITLE Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
JOURNAL Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
REFERENCE Dodg,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
AUTHORS Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

```

```

Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., Mhova,T., Mlenga,V.,
McNeeters,R., Meldrim,J., Meneus,L., Norbu,C., Norman,C.H.,
Murphy,T., Naylor,J., Nguyen,C., O'Neil,D., Oliver,J., Peterson,K.,
O'Connor,T., O'Donnell,P., O'Neil,D., O'Neil,D., O'Neil,D.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rise,C.,
Rise,C., Rogov,P., Roman,J., Roman,J., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougne,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 203690)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collumore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicoll,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J.,
Rise,C., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 24, 2002 this sequence version replaced gi:11875303.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11638
Center clone name: 712 I 15
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190084 bases at least Q40
Consensus quality: 196705 bases at least Q30
Consensus quality: 199369 bases at least Q20
Insert size: 176000; agarose-ep
Insert size: 200390; sum-of-ctngs
Quality coverage: 9.6 in Q20 bases; agarose-ep
Quality coverage: 8.5 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.

```

# TITLE JOURNAL REFERENCE AUTHORS

## TITLE JOURNAL COMMENT

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1  
460: contig of 460 bp in length  
461  
560: gap of 100 bp  
561  
1196: contig of 636 bp in length  
1197  
1296: gap of 100 bp  
1297  
2122: contig of 826 bp in length  
2123  
2222: gap of 100 bp  
2223  
2869: contig of 647 bp in length  
2870  
2969: gap of 100 bp  
2970  
3620: contig of 651 bp in length  
3621  
3720: gap of 100 bp  
3721  
4314: contig of 594 bp in length  
4315  
4414: gap of 100 bp  
4415  
5152: contig of 738 bp in length  
5153  
5252: gap of 100 bp  
5253  
6209: contig of 957 bp in length  
6210  
6309: gap of 100 bp  
6310  
7103: contig of 794 bp in length  
7104  
7203: gap of 100 bp  
7204  
7937: contig of 734 bp in length  
7938  
8037: gap of 100 bp  
8038  
9006: contig of 969 bp in length  
9007  
9106: gap of 100 bp  
9107  
9906: contig of 800 bp in length  
9907  
10006: gap of 100 bp  
10007  
11070: contig of 1064 bp in length  
11071  
11170: gap of 100 bp  
11171  
12297: contig of 1127 bp in length  
12298  
12397: gap of 100 bp  
12398  
13415: contig of 1018 bp in length  
13416  
13515: gap of 100 bp  
13516  
14231: contig of 716 bp in length  
14232  
14331: gap of 100 bp  
14332  
15596: contig of 1265 bp in length  
15597  
15696: gap of 100 bp  
15697  
16422: contig of 766 bp in length  
16423  
16562: gap of 100 bp  
16563  
17742: contig of 1180 bp in length  
17743  
17842: gap of 100 bp  
17843  
18857: contig of 1015 bp in length  
18858  
18957: gap of 100 bp  
18958  
20490: contig of 1533 bp in length  
20491  
20590: gap of 100 bp  
20591  
22210: contig of 1620 bp in length  
22211  
22310: gap of 100 bp  
22311  
23851: contig of 1541 bp in length  
23852  
23951: gap of 100 bp  
23952  
25684: contig of 1733 bp in length  
25685  
25785: gap of 100 bp  
25786  
27006: contig of 1222 bp in length  
27007  
27106: gap of 100 bp  
27107  
28616: contig of 1510 bp in length  
28617  
28716: gap of 100 bp  
28717  
30184: contig of 1468 bp in length  
30185  
30284: gap of 100 bp  
30285  
31303: contig of 1019 bp in length  
31304  
31403: gap of 100 bp  
31404  
32917: contig of 1514 bp in length  
32918  
33017: gap of 100 bp  
33018  
42030: contig of 9013 bp in length  
42031  
42130: gap of 100 bp  
42131  
49012: contig of 6882 bp in length  
49013  
49112: gap of 100 bp  
49113  
62240: contig of 13128 bp in length  
62241  
62340: gap of 100 bp  
62341  
86795: contig of 24455 bp in length  
86796  
86895: gap of 100 bp  
86896  
203690: contig of 116795 bp in length.  
Location/Qualifiers  
1. .203690  
/organism="Homo sapiens"

/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-712115"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .460  
/note="assembly\_fragment"  
clone end:SP6  
vector side:left"  
461..560  
/estimated\_length=100  
561..1196  
/note="assembly\_fragment"  
1197..1296  
/estimated\_length=100  
1297..2122  
/note="assembly\_fragment"  
2123..2222  
/estimated\_length=100  
2223..2869  
/note="assembly\_fragment"  
2870..2969  
/estimated\_length=100  
2970..3620  
/note="assembly\_fragment"  
3621..3720  
/estimated\_length=100  
3721..4314  
/note="assembly\_fragment"  
4315..4414  
Query Match 79.2%; Score 19.8; DB 14; Length 203690;  
Best Local Similarity 91.3%; Pred.No.1.3e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAACTGGAATCTCAGGCTGAG 24  
DB 153635 AGAATCTGGAATCTCAGGCTGAG 153613  
RESULT 30  
AC156131/c  
LOCUS  
DEFINITION Bos taurus clone CH240-47A8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 18  
AC156131 225793 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-47A8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 18  
AC156131  
AC156131.2 GI:68265537  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Bos taurus (cow)  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 225793)  
Muzny, D.M., Metzker, M., Lee, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alshrocks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, O., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Geisregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,



```

/estimated_length=50
204587..204686
/estimated_length=unknown
214264..214441
/estimated_length=178
219210..219309
/estimated_length=unknown
220718..220817
/estimated_length=unknown
222539..222638
/estimated_length=unknown
224292..224391
/estimated_length=unknown

```

## ORIGIN

```

Query Match          79.2%; Score 19.8; DB 14; Length 225793;
Best Local Similarity 91.3%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 AAAAAGTGGAAATCTCAGGCTGAG 24
    ||||| ||||| ||||| ||||| |||||
Db  123076 ACAAAGTGAATTCAGGCTGAG 123054

```

## RESULT 31

```

AC152300/c
LOCUS          227115 bp   DNA   linear   HTG 01-JUL-2005
DEFINITION    Bos taurus clone CH240-1K11, *** SEQUENCING IN PROGRESS ***, 29
               unordered pieces.
ACCESSION     AC152300
VERSION       AC152300.3 GI:68227212
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE        Bos taurus (Cow)

```

```

ORGANISM      Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
               Pecora; Bovidae; Bovinae; Bos.

```

```

1 (bases 1 to 227115)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguitano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huly,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louissegat,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoeleneh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Pruzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

```

```

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajic,D.,
Sned,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabori,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 227115)
Worley,K.C.
Direct Submission
Submitted (01-NOV-2004) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 227115)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 26, 2005 this sequence version replaced gi:58038108.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FAXC
Center clone name: CH240-1K11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 213949 bases at least Q40
Consensus quality: 216851 bases at least Q30
Consensus quality: 219286 bases at least Q20
Estimated insert size: 214671; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3732: contig of 3732 bp in length
* 3733 3782: gap of 50 bp
* 3783 7703: contig of 3921 bp in length
* 7704 7703: gap of unknown length
* 7804 10239: contig of 2436 bp in length
* 10240 10289: gap of 50 bp

```



-----  
 This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one pUC18 subclone; and the assembly was confirmed by restriction digest.  
 -----

Neighboring sequence information:

This clone is overlapped by GSI-25117, GSI-287C17.

-----  
 Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
 -----

FEATURES

source

Location/Qualifiers

1. .234429

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="8"

/map="p21-p11"

/clone="SCB-212E3"

218471. .234429

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="GSI-25117"

/note="overlapping clone"

233972. .234429

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="GSI-287C17"

/note="overlapping clone"

1. .1527

/note="single stranded/single chemistry region"

1752. .1892

/note="single stranded/single chemistry region , pcr product sequence only"

1999. .2209

/note="single stranded/single chemistry region"

2405. .2904

/note="single stranded/single chemistry region"

4045. .4520

/note="single stranded/single chemistry region"

4909. .5465

/note="single stranded/single chemistry region"

5006

/note="low quality region"

5074. .5076

/note="low quality region"

5958. .6158

/note="single stranded/single chemistry region"

8506. .9433

/note="single stranded/single chemistry region"

12839. .12949

/note="single stranded/single chemistry region"

19289. .19505

/note="single stranded/single chemistry region"

22256. .22617

/note="single stranded/single chemistry region"

22915. .23104

/note="single stranded/single chemistry region"

25717. .25784

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

/note="single stranded/single chemistry region"

misc\_feature  
 25979. .26519  
 /note="single stranded/single chemistry region"  
 27780. .27857  
 /note="single stranded/single chemistry region"  
 28946. .28979  
 /note="single stranded/single chemistry region"  
 32258. .32444  
 /note="single stranded/single chemistry region"  
 33909. .34058  
 /note="single stranded/single chemistry region"  
 34482. .34564  
 /note="single stranded/single chemistry region"  
 44437. .44756  
 /note="single stranded/single chemistry region"  
 53510. .53587  
 /note="single stranded/single chemistry region"  
 55018. .55027  
 /note="single stranded/single chemistry region"  
 56892. .57298  
 /note="single stranded/single chemistry region"  
 59734. .59817  
 /note="single stranded/single chemistry region"  
 60277. .60529  
 /note="single stranded/single chemistry region"  
 60712. .60875  
 /note="single stranded/single chemistry region"  
 72531. .72795  
 /note="single stranded/single chemistry region"  
 72810. .72863  
 /note="single stranded/single chemistry region"  
 74251. .74328  
 /note="single stranded/single chemistry region"  
 75828  
 /note="low quality region"  
 76079. .76824  
 /note="single stranded/single chemistry region"  
 81807. .81875  
 /note="single stranded/single chemistry region"  
 83588. .83920  
 /note="single stranded/single chemistry region"  
 86322. .86818  
 /note="single stranded/single chemistry region"  
 88445. .88454  
 /note="single stranded/single chemistry region"  
 89326. .89651  
 /note="single stranded/single chemistry region"  
 90582. .90927  
 /note="single stranded/single chemistry region"  
 95744. .95772  
 /note="single stranded/single chemistry region"  
 98134. .98199  
 /note="single stranded/single chemistry region"  
 98571. .99114  
 /note="single stranded/single chemistry region"  
 102816. .102891  
 /note="single stranded/single chemistry region"  
 102915. .103388  
 /note="single stranded/single chemistry region"  
 103538. .103624  
 /note="single stranded/single chemistry region"  
 103674. .103863  
 /note="single stranded/single chemistry region"  
 108658. .108801  
 /note="single stranded/single chemistry region"  
 110323. .110378  
 /note="single stranded/single chemistry region"  
 114946. .115062  
 /note="single stranded/single chemistry region"  
 115984. .116248  
 /note="single stranded/single chemistry region"  
 118198. .118280  
 /note="single stranded/single chemistry region"  
 119071. .119213  
 /note="single stranded/single chemistry region"



Query Match 79.2%; Score 19.8; DB 8; Length 234429;  
Best Local Similarity 91.3%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAACTGGAATCTCAGGCTGAG 24  
||| ||||| ||||| ||||| |||||  
DB 201360 AAAGACTGGAATCTCAGGCTGAG 201338

RESULT 33  
AC013717/C  
LOCUS AC013717 168813 bp DNA linear PRI 15-APR-2005  
DEFINITION Homo sapiens BAC clone RP11-2415 from 2, complete sequence.  
AC013717  
VERSION AC013717.8 GI:14190753  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 168813)  
AUTHORS Maupin,R. and Edwards,J.  
TITLE The sequence of Homo sapiens BAC clone RP11-2415  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 168813)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 168813)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 168813)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 168813)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (15-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On May 23, 2001 this sequence version replaced gi:13431081.  
-----  
Center: Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@watson.wustl.edu  
-----  
Summary Statistics  
Center project name: H\_NH0024105  
-----

NOTICE:  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see

http://genome.wustl.edu

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-559M23; the clone sequenced to the right is RP11-194L1. Actual start of this clone is at base position 1 of RP11-2415; actual end is at base position 168813 of RP11-2415.

FEATURES  
source

Location/Qualifiers  
1..168813  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-2415"  
/clone\_lib="RPCI-11"  
8128..38413  
/gene="PPM1B"  
join(8128..8987,16152..16269,24909..25020,25422..25479,37355..38413)  
/gene="PPM1B"  
join(8142..8987,16152..16269,24909..25020,25422..25479,37355..37660)  
/gene="PPM1B"  
/note="Homo sapiens protein phosphatase 1B (formerly 2C), magnesium-dependent, beta isoform (PPM1B), transcript variant 1, mRNA.; H\_NH0024105.1  
This gene was based on gi(29558349)  
Continued from H\_NH0559M23.1"

gene

mRNA

CDS

/product="unknown"  
/protein\_id="AA88954.1"  
/db\_xref="GI:62630209"  
/translation="MGAFIDPKTEKHNAGHAGNGRLYGLSSMQGWRVEMEDAHTAVV  
GIPLGLDWSFFAVVDGHSRVANYCSTHLEHTTNEDEFRAGKSGSALELSVENV  
KNGITGTGLKIDVEMRNFSLRNGMDRSGSTAVGMISPKHIYFNCDSRAVLVRNG  
QVCFSTQDKPCNPREKRIQVAGSVMIQVNSGLAYSRALGDYDKVDGKGPTEQ  
LVSPEVPEIILRAEDEFILACDGIWMSNELCYVKSRLVESDLDLNCVNVV  
DTCILKSGRDNMSIVLCFSNAPKVSDEAVKDSLELDKHLSEVEIMEKSGEGMPD  
LAHVMRIILSAENIPLNPPGGLAGKNVIEAVYSLNPHRESDGASDAESGSGKL  
VEALROMRINHNGNYRQLLEMLTSYRLAKVEGESPAEPAATATSSNADAGNPVTMQ  
ESHTESGSLAELDSNEDAGTKWSEKI"  
82402..127740  
/gene="SLC3A1"  
join(82402..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127740)  
/gene="SLC3A1"  
join(82478..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127581)  
/gene="SLC3A1"  
/note="Homo sapiens solute carrier family 3 (cystine,  
dibasic and neutral amino acid transporters, activator of  
cystine, dibasic and neutral amino acid transport), member  
1 (SLC3A1), mRNA.; H\_NH0024105.2  
This gene was based on gi(23308570)"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AA88955.1"  
/db\_xref="GI:62630210"

gene

mRNA

CDS

ESHTESGSLAELDSNEDAGTKWSEKI"  
82402..127740  
/gene="SLC3A1"  
join(82402..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127740)  
/gene="SLC3A1"  
join(82478..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127581)  
/gene="SLC3A1"  
/note="Homo sapiens solute carrier family 3 (cystine,  
dibasic and neutral amino acid transporters, activator of  
cystine, dibasic and neutral amino acid transport), member  
1 (SLC3A1), mRNA.; H\_NH0024105.2  
This gene was based on gi(23308570)"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AA88955.1"  
/db\_xref="GI:62630210"

ESHTESGSLAELDSNEDAGTKWSEKI"  
82402..127740  
/gene="SLC3A1"  
join(82402..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127740)  
/gene="SLC3A1"  
join(82478..82907,87658..87837,88329..88483,92974..93099,  
106913..107032,107945..108069,111085..111280,  
119528..119695,120777..120893,127141..127581)  
/gene="SLC3A1"  
/note="Homo sapiens solute carrier family 3 (cystine,  
dibasic and neutral amino acid transporters, activator of  
cystine, dibasic and neutral amino acid transport), member  
1 (SLC3A1), mRNA.; H\_NH0024105.2  
This gene was based on gi(23308570)"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AA88955.1"  
/db\_xref="GI:62630210"

WITSFYKSLKDRFYGVEDFREDVDFPGTMEFENLVAIAHDKGLKLIIDFIPNHTSD  
KHIFOLSRGTGKTYDVEIYIMHDCITHEGKTIIPNNMLSVYGNSSWHFDEVRNQCFYH  
QWKEQPDILNFRNDQVEIKELRFLWTKGVDFSLDAVKFLAKHLRDEIOVNTK  
QIPDTVTOISELYHDTTTOVGHMDIVRSFRQTMQDYSTPEGRYEMGTAEVAESIDR  
TWYGLPGIADFPFNLSMLDTSVGSNVEITISWNNENPFGKPEWNNMILGSPDS  
SLRTSLGNQYVNMMLLFTLPGTPTTYGEEIGMGNVAANLNESEDYINTLRKSP  
MOWDINSNAGVSEASNTWLPNTSDYHNVQKTPRSALKLYODLSLHANELLIN  
RGWFLHRNDSHVYVTRDLGDIRIFIVLNFGEFTLNLHNMISGLPAKWRILST  
NSADKSGKVDTSIGIFLDKGEGLIFEHTKLLRHQTAFRDRCFVSNRACYSVVLNLY  
TSC"

misc\_feature

93440..93769

/notes="CpG island (%GC=63.6, o/e=0.64, #CpGs=23)"  
/complement(128218..135958)  
/genes="KTA0436"

gene

complement(join(128218..128387,128769..128842,  
129673..129796,130204..130353,133654..133870,  
135879..135958))

mRNA

/genes="KIAA0436"  
complement(join(128298..128387,128769..128842,  
129673..129796,130204..130353,133654..133870,  
135879..135958))

CDS

/notes="Homo sapiens putative L-type neutral amino acid  
transporter, mRNA [cDNA clone MGC:21932 IMAGE:4389731],  
complete cds.; H NH0024105.3  
This gene was based on gi(15341987)"

/codon\_start=1

/product="unknown"

/protein\_id="AA088956.1"

/db\_xref="GI:62630211"

/translations="MDLKNFRPERRVVDGWLAVYVRGGELGLOWHADRLTK  
KLNGLADLRACIKTLHGQFSQSLTLTAFSAGVLAGALNSNPELVAVTLEAPF  
LVLVNTMDTTLTLEELFEWGNPSDEKGNIKYIKRYCPYQNIKPQHPSPHITAYE  
NDRVPLKGVTSKYEALAEHAXDTEGQYTPNIILDIOPGNGHVEDSHKKLTA  
QIKLYEELGLDSTVFDLKKYLF"

ORIGIN

Query Match 77.6%; Score 19.4; DB 8; Length 169813;  
Best Local Similarity 95.2%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 AACGTGAATCTCAGGCTGAGA 25

||||| ||||||| ||||||| ||||||| |||||||

Db 107117 AACTGGTATCTCAGGCTGAGA 107097

RESULT 34

AC073487 54666 bp DNA linear PRI 19-MAR-2003  
LOCUS Homo sapiens 12 BAC RP11-76217 (Robwell Park Cancer Institute Human  
DEFINITION BAC Library) complete sequence.

ACCESSION

AC073487

VERSION

AC073487.34 GI:14578058

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 54666)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshman,F.R., Allen,C.,  
Alsbrooke,S.L., Amaratingue,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavaros,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,  
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,

Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Honsi,P.,  
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.B.,  
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudan,S.,  
Carlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
Lewils,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martin,R.,  
Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,  
Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,  
Minor,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,  
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shim,C.,  
Shooshitari,N., Sieson,I., Sodergren,E., Sonaik,T., Sparks,A.,  
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,  
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,  
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,  
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
Washington,C., Watlington,S., Williams,G., Williamson,A.,  
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
Zorilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 54666)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

3 (bases 1 to 54666)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

4 (bases 1 to 54666)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

5 (bases 1 to 54666)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

On Jun 30, 2001 this sequence version replaced gi:14575758.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST

(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

## Location/Qualifiers

1..54666  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-76217"

repeat\_region

8..292

/rpt\_family="AluSp"

repeat\_region

293..320

/rpt\_family="C(AAAA)n"

repeat\_region

337..454

/rpt\_family="AluJb"

repeat\_region

456..734

/rpt\_family="AluSx"

repeat\_region

complement(1548..1602)

/rpt\_family="L2"

repeat\_region

complement(416..4163)

/rpt\_family="MIR"

repeat\_region

4164..4445

/rpt\_family="AluSx"

repeat\_region

complement(4446..4588)

/rpt\_family="MIR"

repeat\_region

complement(4643..5004)

/rpt\_family="L2"

repeat\_region

5027..5086

/rpt\_family="(TTTTC)n"

repeat\_region

complement(5096..5404)

/rpt\_family="AluSx"

repeat\_region

complement(5738..5799)

/rpt\_family="MIR"

repeat\_region

complement(6208..6518)

/rpt\_family="AluY"

repeat\_region

7220..7355

/rpt\_family="C-rich"

repeat\_region

7583..7801

/rpt\_family="CT-rich"

repeat\_region

7925..7977

/rpt\_family="(CCCG)n"

repeat\_region

8037..8076

/rpt\_family="C-rich"

repeat\_region

8212..8251

/rpt\_family="(CGG)n"

repeat\_region

10383..10412

/rpt\_family="(GGGGA)n"

repeat\_region

10522..10553

/rpt\_family="(TCCCC)n"

repeat\_region

complement(11756..11898)

/rpt\_family="L2"

repeat\_region

14938..15072

/rpt\_family="(CAGAGA)n"

STS

14968..15068

STS  
14988..15146  
/standard\_name="21494"  
repeat\_region  
15260..15281  
/rpt\_family="AT-rich"  
repeat\_region  
16229..16286  
/rpt\_family="MIR"  
STS  
16447..16571  
/standard\_name="79124"  
STS  
16559..16661  
/standard\_name="36508"  
STS  
16927..17081  
/standard\_name="87392"  
repeat\_region  
18299..18606  
/rpt\_family="AluSc"  
repeat\_region  
18612..18788  
/rpt\_family="FAM"  
repeat\_region  
complement(19482..19685)  
/rpt\_family="MIR"  
STS  
21682..21805  
/standard\_name="73682"  
STS  
22161..22330  
/standard\_name="4986"  
STS  
22219..22335  
/standard\_name="14903"  
repeat\_region  
23176..23469  
/rpt\_family="AluSg"  
repeat\_region  
23543..23715  
/rpt\_family="MSTA"  
repeat\_region  
complement(23717..23966)

## Query Match

Best Local Similarity 76.8%; Score 19.2; DB 8; Length 54666;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCGAATCTCAGGCTGAGA 25

Db 13288 AAGAACTGAAAATCAGGCTGAGA 13311

## RESULT 35

AC026926/c

LOCUS

AC026926

DEFINITION

Homo sapiens clone RP11-658L6, LOW-PASS SEQUENCE SAMPLING.

ACCESSION

AC026926.1

VERSION

GI:7328791

KEYWORDS

HTG; HTGS\_PHASE0.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homnidae; Homo.

REFERENCE

1 (bases 1 to 62880)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens, clone RP11-658L6

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 62880)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,B.,

Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Melgrim,J., Meneus,L., Minova,T., Miranda,C., Mienga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

# TITLE JOURNAL

## COMMENT

Direct Submission  
 Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7538

Center clone name: 658\_L\_6

-----

\* NOTE: This record contains 69 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 767: contig of 767 bp in length  
 \* 768 867: gap of 100 bp  
 \* 868 1682: contig of 795 bp in length  
 \* 1663 1762: gap of 100 bp  
 \* 1763 2581: contig of 819 bp in length  
 \* 2582 2681: gap of 100 bp  
 \* 2682 3470: contig of 789 bp in length  
 \* 3471 3570: gap of 100 bp  
 \* 3571 4375: contig of 805 bp in length  
 \* 4376 4475: gap of 100 bp  
 \* 4476 5296: contig of 821 bp in length  
 \* 5297 5396: gap of 100 bp  
 \* 5397 6170: contig of 774 bp in length  
 \* 6171 6270: gap of 100 bp  
 \* 6271 7100: contig of 830 bp in length  
 \* 7101 7200: gap of 100 bp  
 \* 8044: contig of 844 bp in length  
 \* 8045 8144: gap of 100 bp  
 \* 8145 8947: contig of 803 bp in length  
 \* 8948 9047: gap of 100 bp  
 \* 9048 9855: contig of 808 bp in length  
 \* 9856 9955: gap of 100 bp  
 \* 9956 10781: contig of 826 bp in length  
 \* 10782 10881: gap of 100 bp  
 \* 10882 11718: contig of 837 bp in length  
 \* 11719 11818: gap of 100 bp  
 \* 11819 12636: contig of 818 bp in length  
 \* 12637 12736: gap of 100 bp  
 \* 12737 13533: contig of 797 bp in length  
 \* 13534 13633: gap of 100 bp  
 \* 13634 14458: contig of 823 bp in length  
 \* 14457 14556: gap of 100 bp  
 \* 14557 15365: contig of 809 bp in length  
 \* 15366 15465: gap of 100 bp  
 \* 15466 16267: contig of 802 bp in length  
 \* 16268 16367: gap of 100 bp  
 \* 16368 17173: contig of 806 bp in length  
 \* 17174 17273: gap of 100 bp  
 \* 17274 18093: contig of 820 bp in length  
 \* 18094 18193: gap of 100 bp  
 \* 18194 19024: contig of 831 bp in length  
 \* 19025 19124: gap of 100 bp

19125 19277: contig of 803 bp in length  
 \* 19928 20027: gap of 100 bp  
 \* 20028 20860: contig of 833 bp in length  
 \* 20861 20960: gap of 100 bp  
 \* 20961 21688: contig of 728 bp in length  
 \* 21689 21788: gap of 100 bp  
 \* 21789 22620: contig of 832 bp in length  
 \* 22621 22720: gap of 100 bp  
 \* 22721 23530: contig of 810 bp in length  
 \* 23531 23630: gap of 100 bp  
 \* 23631 24468: contig of 838 bp in length  
 \* 24469 24568: gap of 100 bp  
 \* 24569 25366: contig of 798 bp in length  
 \* 25367 25466: gap of 100 bp  
 \* 25467 26282: contig of 816 bp in length  
 \* 26283 26382: gap of 100 bp  
 \* 26383 27204: contig of 822 bp in length  
 \* 27205 27304: gap of 100 bp  
 \* 27305 28131: contig of 827 bp in length  
 \* 28132 28231: gap of 100 bp  
 \* 28232 29065: contig of 834 bp in length  
 \* 29066 29165: gap of 100 bp  
 \* 29166 30004: contig of 839 bp in length  
 \* 30005 30104: gap of 100 bp  
 \* 30105 30941: contig of 837 bp in length  
 \* 30942 31041: gap of 100 bp  
 \* 31042 31835: contig of 794 bp in length  
 \* 31836 31935: gap of 100 bp  
 \* 31936 32761: contig of 826 bp in length  
 \* 32762 32861: gap of 100 bp  
 \* 32862 33694: contig of 833 bp in length  
 \* 33695 33794: gap of 100 bp  
 \* 33795 34602: contig of 808 bp in length  
 \* 34603 34702: gap of 100 bp  
 \* 34703 35527: contig of 825 bp in length  
 \* 35528 35627: gap of 100 bp  
 \* 35628 36442: contig of 815 bp in length  
 \* 36443 36542: gap of 100 bp  
 \* 36543 37376: contig of 834 bp in length  
 \* 37377 37476: gap of 100 bp  
 \* 37477 38264: contig of 788 bp in length  
 \* 38265 38364: gap of 100 bp  
 \* 38365 39135: contig of 771 bp in length  
 \* 39136 39235: gap of 100 bp  
 \* 39236 40045: contig of 810 bp in length  
 \* 40046 40145: gap of 100 bp  
 \* 40146 40945: contig of 800 bp in length  
 \* 40946 41045: gap of 100 bp  
 \* 41046 41762: contig of 717 bp in length  
 \* 41763 41862: gap of 100 bp  
 \* 41863 42696: contig of 834 bp in length  
 \* 42697 42796: gap of 100 bp  
 \* 42797 43635: contig of 839 bp in length  
 \* 43636 43735: gap of 100 bp  
 \* 43736 44530: contig of 795 bp in length  
 \* 44531 44630: gap of 100 bp  
 \* 44631 45456: contig of 826 bp in length  
 \* 45457 45556: gap of 100 bp  
 \* 45557 46382: contig of 826 bp in length  
 \* 46383 46482: gap of 100 bp  
 \* 46483 47297: contig of 815 bp in length  
 \* 47298 47397: gap of 100 bp  
 \* 47398 48210: contig of 813 bp in length  
 \* 48211 48310: gap of 100 bp  
 \* 48311 49152: contig of 842 bp in length  
 \* 49153 49252: gap of 100 bp  
 \* 49253 50063: contig of 811 bp in length  
 \* 50064 50163: gap of 100 bp  
 \* 50164 50984: contig of 821 bp in length  
 \* 50985 51084: gap of 100 bp  
 \* 51085 51922: contig of 838 bp in length  
 \* 51923 52022: gap of 100 bp  
 \* 52023 52841: contig of 819 bp in length

```

* 52842 52941: gap of 100 bp
* 52942 53770: contig of 829 bp in length
* 53771 53870: gap of 100 bp
* 53871 54648: contig of 778 bp in length
* 54649 54748: gap of 100 bp
* 54749 55542: contig of 794 bp in length
* 55543 55642: gap of 100 bp
* 55643 56466: contig of 824 bp in length
* 56467 56566: gap of 100 bp
* 56567 57401: contig of 835 bp in length
* 57402 57501: gap of 100 bp
* 57502 58319: contig of 818 bp in length
* 58320 58419: gap of 100 bp
* 58420 59234: contig of 815 bp in length
* 59235 59334: gap of 100 bp
* 59335 60165: contig of 831 bp in length
* 60166 60265: gap of 100 bp
* 60266 61081: contig of 816 bp in length
* 61082 61181: gap of 100 bp
* 61182 61969: contig of 788 bp in length
* 61970 62070: contig of 811 bp in length.
* 62070 62880: contig of 811 bp in length.

```

## FEATURES

```

Location/Qualifiers
Query Match 76.8%; Score 19.2; DB 14; Length 62880;
Best Local Similarity 87.5%; Pred.No.2.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 AAAAACTGGAACTCTCAGGCTGAG 24

Db 26104 AAAAACTGGAACTCTCAGGCTGAG 26081

## RESULT 36

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: http://www.sanger.ac.uk/Projects/D\_rerio/faq.shtml#dataeight On Jun 25, 2003 this sequence version replaced gi:32169125.

----- Genome Center

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk

-----

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

CH211-272H10 is from a CHORI-211 BAC library  
VECTOR: pTARBAC2.1

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

Any regions longer than 1kb tagged as misc-feature 'unsure' are part of a tandem repeat of more than 10kb in length where it has not been possible to anchor the base differences between repeat copies. The region has been built up based on the repeat element to match the total size of repeat indicated by restriction digest, but repeat copies may not be in the correct order and the usual finishing criteria may not apply.

## FEATURES

source

Location/Qualifiers

1..99176

/organism="Danio rerio"

/mol\_type="genomic DNA"

/db\_xref="taxon:7955"

/chromosome="4"

/clone="CH211-272H10"

/clone\_lib="CHORI-211"

2000

/note="Clone\_left\_end: CH211-11C20"

23138..23184

/note="Single clone region. Assembly confirmed by

restriction digest data."

36912..36917

36941

complement(62236..71148)

/gene="btg1"

/locus\_tag="CH211-272H10.1-001"

complement(join(62236..65494,70781..71148))

/gene="btg1"

/locus\_tag="CH211-272H10.1-001"

/product="B-cell translocation gene 1"

/note="match: CDNAs: Em:BC056691.1

match: ESTs: Em:AW175489.1 Em:AW421052.1 Em:BE201884.1

Em:BE557354.1 Em:BE605378.1 Em:BE605928.1 Em:BE606028.1

Em:BI881139.1 Em:BI888289.1 Em:BI891235.1 Em:BI891276.1

Em:BI891276.1 Em:BI891276.1 Em:BI891276.1 Em:BI891276.1

Em:CA474442.1 Em:CA474859.1 Em:CA496493.1 Em:CA496493.1

Em:CD283714.1 Em:CD587119.1 Em:CD59529.1 Em:CD59547.1

Em:CD778016.1 Em:CF103301.1 Em:CF924881.1 Em:CF996149.1

Em:CK015896.1 Em:CK016458.1 Em:CK018724.1 Em:CK024547.1

Em:CK236966.1 Em:CK238220.1 Em:CK339087.1 Em:CK354840.1

Em:CK361434.1 Em:CK362051.1 Em:CK397124.1 Em:CK401068.1

complement(join(65091..65494,70781..70925))

/gene="btg1"

/locus\_tag="CH211-272H10.1-001"

/standard\_name="OTTDARP0000007966"

/note="match: proteins: Sw:P31607 Sw:P34743 Sw:P53348

Sw:Q63073 Tr:AAH56691"

/codon\_start=1

/product="B-cell translocation gene 1"

/protein\_id="CAH68953.1"

/db\_xref="GI:55251104"

/translation="WHITLCARGTWKPEINAAVGLSLRFLRIKHVNDRLQTFQSOTLQ

DILAQYKHFWFPDRKGRIRINRKHMDPLVGGQGRIGLSIQYLILPSEL

LWVDPFVSYRIGEDSGICVLYESHGPTNGNPSTTGNISPASSVTVQVSAWVESHIS

KEELLVLGRTSPAKPYMVTSS"

99176

/note="Clone\_left\_end: CH211-272H10"

## ORIGIN

```

Query Match 76.8%; Score 19.2; DB 5; Length 99176;
Best Local Similarity 87.5%; Pred.No.2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAACTCTCAGGCTGAGA 25
    |||||
Db 70252 AAAAACTGGAACTCCAGCTTTAGA 70275
    |||||

```

RESULT 37  
AC090678/c  
LOCUS  
DEFINITION  
AC090678 Homo sapiens chromosome 12 clone RP11-412M2, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 31 unordered pieces.  
AC090678  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC090678 105736 bp DNA linear HTG 08-JAN-2003  
Homo sapiens chromosome 12 clone RP11-412M2, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 31 unordered pieces.  
AC090678  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC090678.3 GI:21431036  
HTG; HTGS\_PHASE1.  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 105736)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Anarstunge, H.C., Are, J.R., Ayale, M., Banks, I.,  
Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
Bouchar, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J.J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kreshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louleeged, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, B., Nwokenko, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peeters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villaion, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 105736)  
Worley, K.C.  
Direct Submission  
Submitted (08-MAR-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 105736)  
Worley, K.C.  
Direct Submission  
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 17, 2002 this sequence version replaced gi:14547649.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HCQ1  
Center clone name: RP11-412M2  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 131977 bases at least Q40  
Consensus quality: 14820 bases at least Q30  
Consensus quality: 150638 bases at least Q20  
Estimated insert size: 3481; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 2113: contig of 2113 bp in length  
\* 2114 2213: gap of unknown length  
\* 2214 4755: contig of 2362 bp in length  
\* 4576 4675: gap of unknown length  
\* 4676 6855: contig of 2180 bp in length  
\* 6856 9483: gap of unknown length  
\* 9484 9583: contig of 2528 bp in length  
\* 9584 11332: gap of unknown length  
\* 11333 11733: gap of unknown length  
\* 11733 13905: contig of 2173 bp in length  
\* 13906 14005: gap of unknown length  
\* 14006 16168: contig of 2063 bp in length  
\* 16169 18363: gap of unknown length  
\* 18364 18463: gap of unknown length  
\* 18464 20771: contig of 2208 bp in length  
\* 20772 22976: contig of 2205 bp in length  
\* 22977 23076: gap of unknown length  
\* 23077 25231: contig of 2154 bp in length  
\* 25232 28408: gap of unknown length  
\* 28409 31188: contig of 3078 bp in length  
\* 31189 31288: gap of unknown length  
\* 31289 33857: contig of 2680 bp in length  
\* 33858 33957: gap of unknown length  
\* 33958 35993: contig of 2036 bp in length  
\* 35994 40174: gap of unknown length  
\* 40175 40274: contig of 4081 bp in length  
\* 40275 43333: gap of unknown length  
\* 43334 43433: contig of 3059 bp in length  
\* 43434 45838: gap of unknown length  
\* 45839 45938: contig of 2405 bp in length  
\* 45939 48604: gap of unknown length  
\* 48605 48704: contig of 2666 bp in length  
\* 48705 51037: gap of unknown length  
\* 51038 51137: contig of 2333 bp in length  
\* 51138 54665: gap of unknown length  
\* 54666 54765: contig of 3528 bp in length  
\* 54766 58612: gap of unknown length  
\* 58613 58712: contig of 3847 bp in length  
\* 58713 62606: contig of 3894 bp in length  
\* 62607 67021: gap of unknown length  
\* 67022 67021: contig of 4315 bp in length  
\* 67022 67021: contig of 4315 bp in length





Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc>.  
Drafting Center Code: WIBR  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu).

FEATURES SOURCE

Source	1. .152222	/organism="Mus musculus"
	.	/mol_type="genomic DNA"
	/db_xref="taxon:10090"	/chromosomes="10"
	/clone="RP24-212P21"	
repeat_region	1921. .2050	/rpt_family="(CA) n"
repeat_region	complement(2521. .2592)	
repeat_region	3555. .3820	/rpt_family="BC1_Mm"
repeat_region		/rpt_family="ORR1D"
repeat_region	3822. .3929	/rpt_family="(TATG)n"
repeat_region	complement(4101. .4171)	
repeat_region		/rpt_family="MER5A"
repeat_region	complement(4482. .4646)	
repeat_region		/rpt_family="B3A"
repeat_region	complement(4837. .11424)	
repeat_region	/rpt_family="LIMD_A"	
repeat_region	12061. .12090	/rpt_family="(TTAA) n"
repeat_region	complement(12091. .12176)	
repeat_region		/rpt_family="PB1D9"
repeat_region	12207. .12228	
repeat_region	/rpt_family="AT_rich"	
repeat_region	12911. .13131	
repeat_region		/rpt_family="MLT1A0"
repeat_region	13372. .13371	
misc_feature	/rpt_family="C-rich"	
	13295	/notes="C / A"
misc_feature		/function="unsure"
	13299	
	/notes="T / A"	
	/function="unsure"	
repeat_region	13366. .13538	
repeat_region	/rpt_family="CM-rich"	
repeat_region	13566. .13634	
repeat_region	/rpt_family="ID4"	
repeat_region	complement(14772. .14889)	
repeat_region	/rpt_family="PB1"	
repeat_region	14926. .14953	
repeat_region	/rpt_family="(TTAA) n"	
repeat_region	complement(14970. .15117)	
repeat_region	/rpt_family="RSINE1"	
repeat_region	15143. .15294	
repeat_region	/rpt_family="(CTCA) n"	
repeat_region	complement(15556. .16428)	
repeat_region	/rpt_family="RMR17D"	
repeat_region	16447. .16474	

QY	2	AAAACTGGAAATCTCAGGCTGAGA	25
DB	82371	AAAAACTGCAATCTCAGGCTTAAA	82348
RESULT 39	AC158544	152222 bp	DNA linear ROD 17-JUN-2005
LOCUS	Mus musculus 10 BAC RP24-212P21	(Roswell Park Cancer Institute	
DEFINITION	(C57BL/6J Male) Mouse BAC Library)	complete sequence.	
ACCSSION	AC158544	AC105988	
VERSION	AC158544.1	GI:61675713	
SOURCE	HTG.		
ORGANISM	Mus musculus (house mouse)		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;			
Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 152222)		
AUTHORS	Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P., Arredondo,H., Bandaranaike,D., Bangura,L.H., Beltran,B., Beltran,R.P., Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Bubay,C., Burch,P., Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I., Chacko,J., Chanrour,M., Chavez,D., Chen,A., Cheng,G., Chen,R., Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A., Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Delgado,O., Denson,S., Detamo,C., Ding,Y., Dinh,H., Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K., Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M., Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P., Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T., Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D., Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haeberlen,K., Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P., Hawes,A., Hawkins,B., Hayes,S., Hemphill,L., Hernandez,J., Hines,S., Hitchens,M., Hodgson,A., Hogues,M., Hollins,B., Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L., Jacobs,T., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S., Keys,L., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F., Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B., Li,Y., Li,Z., Linell,M., Liu,W., Liu,X.-S., Liu,Y., Liyangde,D., London,P., Lopez,J., Lorensuhea,L., Lozado,R., Luk,T., Madu,R., Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E., McCrelland,H., McPherson,J., Mercadao,C., Metzker,M., Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M., Murray,D., Nazareth,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E., Nott,A., Nwaokemelehu,O., Obregon,M., Ochi-Okorie,C., Odhe,E., Okuwonu,G., Okuwonu,K., Parker,D., Pasternak,S., Patel,B., Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T., Primus,E., Pu,L.-L., Puzao,M., Qin,X., Quinn,A., Quiroz,J., Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S., Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W., Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y., Siesson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R., Trejos,Z., Usmani,K., Vargo,C., Verduzo,D., Villasana,D., Virk,D., Volkov,A., Waldron,L., Walker,B., Wang,Q.O., Wang,S., Warren,J., Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R., Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L., Zhang,Z., Zhou,J., Weinstein,G. and Gibbs.R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 152222)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 152222)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-JUN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	On Mar 22, 2005, this sequence version replaced gi:45917490.		

repeat_region	/rpt family="TCCC)n"	16482. .16657
repeat_region	/rpt family="(TTTC)n"	complement(16675. .16754)
repeat_region	/rpt family="ID4"	17131. .17233
repeat_region	/rpt family="MTE"	17994. .18569
repeat_region	/rpt family="MT2B2"	complement(19676. .19951)
repeat_region	/rpt family="B4A"	20123. .20493
repeat_region	/rpt family="ORR1b"	complement(20976. .21179)
repeat_region	/rpt family="B3"	21351. .21371
repeat_region	/rpt family="AT rich"	complement(21917. .22093)
repeat_region	/rpt family="B4A"	complement(22154. .22282)
repeat_region	/rpt family="L2"	complement(23052. .23145)
repeat_region	/rpt family="L2"	23823. .23907
repeat_region	/rpt family="CT-rich"	complement(24930. .25155)
repeat_region	/rpt family="ID_B1"	25258. .25437
repeat_region	/rpt family="(TTTC)n"	complement(25439. .25590)
repeat_region	/rpt family="Lx6"	26605. .26651
repeat_region	/rpt family="(TC)n"	complement(26931. .27291)
repeat_region	/rpt family="ORR1D"	complement(27369. .27406)
repeat_region	/rpt family="RMER1A"	complement(27435. .27671)
repeat_region	/rpt family="RMERIC"	27645. .27749
repeat_region	/rpt family="RMERIC"	27839. .27937
repeat_region	/rpt family="MIR"	complement(29389. .30572)
repeat_region	/rpt family="RMER12"	30984. .31173
repeat_region	/rpt family="B2_Mm2"	31176. .31298
repeat_region	/rpt family="(GGA)n"	31945. .32172
repeat_region	/rpt family="MTB_Mm"	32176. .32197
repeat_region	/rpt family="AT rich"	32594. .32656
repeat_region	/rpt family="MTB_Mm"	32728. .32917
repeat_region	/rpt family="B3"	33223. .33317
repeat_region	/rpt family="B1F"	33286. .33467
STS	/standard_name="MHAa67g10.seq"	

Query Match 76.8%; Score 19.2; DB 9; Length 152222;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGAGA 25  
|||||  
DB 83238 AAAAAGTGAATCTCAGGCTGAGA 83215  
|||||

RESULT 40  
AC104037/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC104037 Homo sapiens chromosome 8, clone RP11-692P18, complete sequence.  
AC104037  
AC104037.3 GI:19352312  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 152267)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 8, clone RP11-692P18  
Unpublished  
2 (bases 1 to 152267)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 152267)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (12-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2002 this sequence version replaced gi:18377276.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L22011  
 Center clone name: 692\_P\_18

## FEATURES

## source

```

1. .152267
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="8"
   /map="8"
   /clone="RP11-692P18"
   /clone_lib="RPC1-11 Human Male BAC"
2. .440
   /rpt_family="L1MD1"
35. .846
   /rpt_family="L1MD1"
complement(2002..2456)
   /rpt_family="MER115"
3150..3170
   /rpt_family="AT-rich"
3204..6714
   /rpt_family="L1PA10"
6767..6982
   /rpt_family="L2"
complement(6994..7071)
   /rpt_family="Tigger7"
7072..7179
   /rpt_family="MER47A"
complement(7181..7309)
   /rpt_family="L1PA16"
7319..7881
   /rpt_family="L1PB2"
complement(7892..9413)
   /rpt_family="L1PA16"
9414..9662
   /rpt_family="MER47A"
complement(9677..9938)
   /rpt_family="Tigger7"
9942..9973
   /rpt_family="L2"
complement(9969..10388)
   /rpt_family="L2"
complement(10394..10453)
   /rpt_family="MIR"
complement(10653..10957)
   /rpt_family="AluSx"
11200..11226
   /rpt_family="(A)n"
11402..11666
   /rpt_family="LTR16A1"
14862..14892
   /rpt_family="AT-rich"
complement(15052..15120)
   /rpt_family="MLT1F1"
complement(15123..15398)
   /rpt_family="AluSx"
15445..15688
   /rpt_family="Aluub"
15698..15876
   /rpt_family="MIR"
16774..17292
   /rpt_family="MLT1F1"
17493..19099
   /rpt_family="L1PA13"
19820..20096
   /rpt_family="L2"
20267..20552
   /rpt_family="Aluub"
20557..20991
   /rpt_family="L2"

```

```

repeat_region 21007..21277
   /rpt_family="AluSg"
complement(21924..22230)
   /rpt_family="AluSg"
complement(23277..23569)
   /rpt_family="AluSx"
complement(23872..23913)
   /rpt_family="MIR3"
23956..24031
   /rpt_family="L1MCS"
complement(24752..25065)
   /rpt_family="L2"
complement(25942..26173)
   /rpt_family="AluJo"
complement(27746..27906)
   /rpt_family="MER112"
27919..28057
   /rpt_family="MIR"
complement(28222..28337)
   /rpt_family="L2"
complement(29041..29436)
   /rpt_family="L1PA5"
29442..29626
   /rpt_family="L1PA5"
30000..30208
   /rpt_family="MER58A"
30932..30996
   /rpt_family="GA-rich"
31046..31399
   /rpt_family="MLT1C"
complement(31424..32306)
   /rpt_family="L1MA7"
complement(32307..33165)
   /rpt_family="L1"
complement(33166..33454)
   /rpt_family="AluSx"
complement(33455..33804)
   /rpt_family="L1"
33805..34087
   /rpt_family="AluJb"
complement(34088..34296)
   /rpt_family="L1"
complement(34334..35294)
   /rpt_family="L1"
complement(35305..35452)
   /rpt_family="MLT1B"
complement(35556..36179)
   /rpt_family="L1PA4"
complement(36188..36459)
   /rpt_family="L1M2"
complement(37672..37816)
   /rpt_family="L1M2"
37830..37940
   /rpt_family="MLT1C"

```

Query Match 76.8%; Score 19.2; DB 8; Length 152267;  
 Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGTGAATCTCAGGCTGAG 24  
 |||||  
 Db 43430 AAGAAGTGAATCTCAGTCTGAG 43407

## RESULT 41

AL671979/c

LOCUS

DEFINITION

AL671979

sequence.

ACCESSION

AL671979.5

VERSION

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

AL671979

Mouse DNA sequence from clone RP23-314B7 on chromosome X, complete

154170 bp

DNA

linear

ROD 05-APR-2002



```
repeat_region /rpt family="L1"
repeat_region 10149. .10371
repeat_region /rpt family="MER1_type"
repeat_region 10372. .10451
repeat_region /rpt family="L1"
repeat_region 10554. .10668
repeat_region /rpt family="L1"
repeat_region 10666. .10985
repeat_region /rpt family="L1"
repeat_region 11521. .11701
repeat_region /rpt family="B2"
repeat_region 12326. .12382
repeat_region /rpt family="L1"
repeat_region 12385. .12477
repeat_region /rpt family="L1"
repeat_region 12721. .12859
repeat_region /rpt family="Alu"
repeat_region 13133. .13248
repeat_region /rpt family="L1"
repeat_region 14350. .14424
repeat_region /rpt family="ID"
repeat_region 14461. .14802
repeat_region /rpt family="ERVk"
repeat_region 15333. .15647
repeat_region /rpt family="ERVl"
repeat_region 16395. .16565
repeat_region /rpt family="MER1_type"
repeat_region 17745. .17857
repeat_region /rpt family="L1"
repeat_region 17858. .18024
repeat_region /rpt family="MER1_type"
repeat_region 18788. .19010
repeat_region /rpt family="B4"
repeat_region 19646. .19889
repeat_region /rpt family="L1"
repeat_region 20866. .21180
repeat_region /rpt family="L1"
repeat_region 21233. .21325
repeat_region /rpt family="L1"
repeat_region 22395. .22602
repeat_region /rpt family="MER1_type"
repeat_region 23561. .23611
repeat_region /rpt family="ERV1"
repeat_region 23752. .23883
repeat_region /rpt family="B4"
repeat_region 23936. .24123
repeat_region /rpt family="ERV1"
repeat_region 24565. .24632
repeat_region /rpt family="ERVl"
repeat_region 24985. .25114
repeat_region /rpt family="ERVl"
repeat_region 25279. .25584
repeat_region /rpt family="B4"
repeat_region 27940. .28070
repeat_region /rpt family="B2"
repeat_region 29585. .29941
repeat_region /rpt family="MaLR"
repeat_region 30046. .30171
repeat_region /rpt family="L1"
repeat_region 31130. .31340
repeat_region /rpt family="B2"
repeat_region 31567. .31624
repeat_region /rpt family="ERV1"
repeat_region 31625. .31757
repeat_region /rpt family="Alu"
repeat_region 31847. .31930
repeat_region /rpt family="L1"
repeat_region 34010. .34213
repeat_region /rpt family="B2"
repeat_region 34530. .34645
repeat_region /rpt family="Alu"
repeat_region 34813. .34932
repeat_region /rpt family="Alu"
```

```
repeat_region 35001. .35428
repeat_region /rpt family="MaLR"
repeat_region 35519. .35874
repeat_region /rpt family="MaLR"
repeat_region 37067. .37511
repeat_region /rpt family="MaLR"
repeat_region 38737. .38875
repeat_region /rpt family="B4"
repeat_region 41238. .41514
repeat_region /rpt family="ERV1"
repeat_region 42758. .43087
repeat_region /rpt family="MaLR"
repeat_region 43088. .43195
repeat_region /rpt family="MaLR"
repeat_region 43212. .43331
repeat_region /rpt family="L1"
repeat_region 45072. .45495
repeat_region /rpt family="MaLR"
repeat_region 45560. .45691
repeat_region /rpt family="Alu"
repeat_region 45823. .46006
repeat_region /rpt family="B2"
repeat_region 46221. .46618
repeat_region /rpt family="MaLR"
repeat_region 47277. .47344
repeat_region /rpt family="ID"
repeat_region 47782. .48147
repeat_region /rpt family="L1"
repeat_region 49447. .49505
repeat_region /rpt family="L1"
repeat_region 49526. .49709
repeat_region /rpt family="L1"
repeat_region 50291. .50443
repeat_region /rpt family="MaLR"
repeat_region 51494. .51898
```

Query Match 76.8%; Score 19.2; DB 9; Length 155268;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 AAAAATGGAAATCTCAGGCTGAGA 25  
|||||  
Db 33162 AAAAATGGAAATCTCAGGCTGAGA 33139

RESULT 43  
AF307158/c

LOCUS AF307158 155669 bp DNA linear HTG 26-JUL-2002  
DEFINITION Homo sapiens chromosome 8 clone RP11-117P17 map 8p12, WORKING DRAFT  
SEQUENCE, 8 unordered pieces.

ACCESSION AF307158.2 GI:14389408  
VERSION AF307158.2  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 155669)

AUTHORS Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,  
Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,  
Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.

Chromosome 8 genomic sequence

Unpublished

REFERENCE 2 (bases 1 to 155669)

AUTHORS Genome Sequencing Center Jena.

Direct Submission

TITLE Submitted (21-SEP-2000) Genome Analysis, Institute of Molecular

JOURNAL Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

On Jun 13, 2001 this sequence version replaced gi:10334826.

COMMENT ----- Genome Center

Center: Institute of Molecular Biotechnoloy

Center code: IMB

web site: <http://genome.imb-jena.de/>  
 Contact: gscj-submitgenome.imb-jena.de  
 ----- Project Information  
 Center project name: H425  
 Center clone name: RP11-117P17

**Sequence Quality Assessment:**

**Sequence Quality Assessment:**  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality 10. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

- \* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	8570:	contig of 8570 bp in length
*	8571	8670:	gap of unknown length
*	8671	36041:	contig of 27371 bp in length
*	36042	36141:	gap of unknown length
*	53289	53289:	contig of 17148 bp in length
*	53290	53389:	gap of unknown length
*	53390	80392:	contig of 27003 bp in length
*	80393	80492:	gap of unknown length
*	80943	109845:	contig of 29353 bp in length
*	109846	109945:	gap of unknown length
*	109946	126299:	contig of 16354 bp in length
*	126300	126399:	gap of unknown length
*	126400	136366:	contig of 9967 bp in length
*	136367	136466:	gap of unknown length
*	136467	155669:	contig of 19203 bp in length

## Location/Qualifiers

```
1. 155669
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p12"
/clone="RP11-117p17"
```

8571, .8670 /estimated length=unknown  
36042, .36141 /estimated length=unknown  
53290, .53389 /estimated length=unknown  
80393, .80492 /estimated length=unknown  
109846, .109945 /estimated length=unknown  
126300, .126399 /estimated length=unknown  
136367, .136466 /estimated length=unknown

## ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 155669;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;

[illegible]

**NOTICE:**

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC140243.

## FEATURES

## source

1. 162528

Location/Qualifiers

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="9"

/map="9"

/clone="RP24-296C19"

/clone\_lib="RPC1-24"

932. 1076

/rpt\_family="Alu"

1787. 1971

/rpt\_family="MIR"

2617. 2756

/rpt\_family="MIR"

5649. 5775

/rpt\_family="Alu"

5776. 5857

/rpt\_family="B4"

7268. 7672

/rpt\_family="RMR17C"

8436. 8519

/rpt\_family="MIR"

8689. 8757

/rpt\_family="L2"

8817. 8949

/rpt\_family="Alu"

9301. 9660

/rpt\_family="MaLR"

10124. 10303

/rpt\_family="B2"

10355. 10496

/rpt\_family="B4"

10571. 10718

/rpt\_family="Alu"

10779. 10908

/rpt\_family="MaLR"

10906. 11027

/rpt\_family="B2"

11070. 11222

/rpt\_family="Alu"

11789. 12045

/rpt\_family="B4"

12241. 12358

/rpt\_family="B2"

12377. 12536

/rpt\_family="B4"

12593. 12650

/rpt\_family="L1"

12655. 12800

/rpt\_family="Alu"

12806. 12922

/rpt\_family="L1"

13137. 13281

/rpt\_family="Alu"

13289. 13360

/rpt\_family="ID"

13532. 13720

/rpt\_family="B2"

14411. 14550

/rpt\_family="Alu"

17393. 17539

/rpt\_family="Alu"

17784. 17977

/rpt\_family="B2"

18370. 18480

/rpt\_family="B4"

repeat\_region

18485. 18614

/rpt\_family="Alu"

repeat\_region

19195. 19364

/rpt\_family="B2"

tRNA

19201. 19273

/product="tRNA-Ser"

/note="Likely pseudogene (HMM Sc=39.82 / Sec struct

Sc=-11.21)"

repeat\_region

19964. 20105

/rpt\_family="Alu"

repeat\_region

20471. 20688

/rpt\_family="B2"

repeat\_region

20770. 20868

/rpt\_family="Alu"

repeat\_region

20960. 21086

/rpt\_family="MIR"

repeat\_region

21223. 21415

/rpt\_family="B2"

repeat\_region

21600. 21746

/rpt\_family="Alu"

repeat\_region

21991. 22060

/rpt\_family="ID"

repeat\_region

22845. 22953

/rpt\_family="Alu"

repeat\_region

22965. 23123

/rpt\_family="B2"

repeat\_region

23145. 23338

/rpt\_family="B2"

repeat\_region

23760. 23895

/rpt\_family="B2"

repeat\_region

23899. 24191

/rpt\_family="MaLR"

repeat\_region

24476. 24585

/rpt\_family="B2"

repeat\_region

24672. 24818

/rpt\_family="Alu"

repeat\_region

25228. 25408

/rpt\_family="B2"

repeat\_region

25503. 25688

/rpt\_family="B2"

repeat\_region

25796. 25880

/rpt\_family="Alu"

repeat\_region

25986. 26135

/rpt\_family="Alu"

repeat\_region

26223. 26740

/rpt\_family="L1"

repeat\_region

26750. 26900

/rpt\_family="Alu"

repeat\_region

29766. 29872

/rpt\_family="Alu"

repeat\_region

30031. 30314

/rpt\_family="B4"

repeat\_region

31490. 31578

/rpt\_family="Alu"

repeat\_region

32188. 32267

/rpt\_family="ID"

repeat\_region

32299. 32426

/rpt\_family="B4"

repeat\_region

32479. 32676

/rpt\_family="B4"

repeat\_region

32909. 33004

/rpt\_family="L1"

repeat\_region

33598. 33685

Query Match 76.8%; Score 19.2; DB 9; Length 162528;

Best Local Similarity 87.5%; Pred. No. 2.6e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGAATCTCAGCTGAG 24

|||||

Db 15606 AAAAACTGAATCTCAGCTGAG 15583



RESULT 45  
AC119551  
LOCUS  
TITLE  
Rattus norvegicus clone CH230-412B6, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
AC119551  
AC119551.5 GI:25008679  
VERSION  
HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Rattus;  
1 (bases 1 to 164792)  
REFERENCE  
AUTHORS  
Muzny, D., Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guvarra, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, J., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Loresushewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Manthey, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, W., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 164792)  
Worley, K. C.  
Direct Submission  
TITLE  
Submitted (28-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 164792)  
REFERENCE  
AUTHORS  
Rat Genome Sequencing Consortium.

TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23808213.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWFF  
Center clone name: CH230-412B6  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 152401 bases at least Q40  
Consensus quality: 154698 bases at least Q30  
Consensus quality: 155987 bases at least Q20  
Estimated insert size: 154442; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 164792: contig of 164792 bp in length.  
----- Location/Qualifiers  
1. 164792  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-412B6"  
1. 1456  
/note="wgs\_end\_extension  
clone\_end:77"  
5329..6087  
/note="clone\_boundary  
clone\_end:77"  
site:  
end\_sequence:BZ142077"  
complement(159147..160072)  
/note="clone\_boundary  
clone\_end:Sp6"  
site:  
end\_sequence:BZ142078"  
161480..162593  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
162644..164792  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
-----  
FEATURES  
source  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 164792;

Best Local Similarity 87.5%; Pred. No. 2.6e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAACTGGATCTCAGGCTGAG 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 15404 AAAAAAATAGATCTCAGGCTGAG 15427

RESULT 46  
 AC132235/c  
 LOCUS AC132235 180462 bp DNA linear ROD 15-MAY-2004  
 DEFINITION Mus musculus BAC clone RP24-497H15 from chromosome 17, complete  
 sequence.  
 AC132235  
 AC132235.3 GI:46559523  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 180462)  
 VanBrunt, A., Cotton, M., Van Brunt, A., Kozlowski, A. and Haakenson, W.  
 TITLE The sequence of Mus musculus BAC clone RP24-497H15  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 180462)  
 McPherson, J.D. and Waterston, R.H.  
 AUTHORS Direct Submission  
 TITLE Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 180462)  
 Wilson, R.K.  
 AUTHORS Direct Submission  
 TITLE Submitted (05-MAR-2004) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 180462)  
 Wilson, R.K.  
 AUTHORS Direct Submission  
 TITLE Submitted (24-APR-2004) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 180462)  
 Wilson, R.K.  
 AUTHORS Direct Submission  
 TITLE Submitted (15-MAY-2004) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA  
 COMMENT On Apr 24, 2004 this sequence version replaced gi:45120391.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@watson.wustl.edu  
 ----- Summary Statistics  
 Center project name: M\_BB0497H15  
 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone, fosmid clone or direct clone walk sequence.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to obtain the consensus sequence; and  
 the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. Wes Warren,  
 Department of Genetics, Washington University, St. Louis MO. For  
 additional information about the map position of this sequence, see  
 http://genome.wustl.edu

## SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and  
 coworkers (http://www.chori.org) from male C57BL/6J mouse spleen  
 and/or brain genomic DNA. The clone and detailed information can be  
 obtained from Pieter de Jong and coworkers at http://www.chori.org

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is  
 overlapped by AC133946 and AL592112.

FEATURES	Location/Qualifiers
source	1..180462 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="17" /map="17"
repeat_region	/clone="RP24-497H15" /clone_lib="RPCI-24" 2012..2086 /rpt_family="MIR"
repeat_region	4862..5007 /rpt_family="B2"
repeat_region	5026..5102 /rpt_family="Alu"
repeat_region	5156..5427 /rpt_family="B4"
repeat_region	7815..7914 /rpt_family="MER1_type"
repeat_region	8205..8351 /rpt_family="Alu"
repeat_region	9343..9404 /rpt_family="ERV1"
repeat_region	9657..9838 /rpt_family="ERV1"
repeat_region	10281..10453 /rpt_family="B2"
repeat_region	11584..11807 /rpt_family="ERVK"
repeat_region	12089..12544 /rpt_family="L1"
repeat_region	12621..12693 /rpt_family="MaLR"
repeat_region	13372..13642 /rpt_family="MaLR"
repeat_region	13643..14039 /rpt_family="MaLR"
repeat_region	14040..14155 /rpt_family="MaLR"
repeat_region	14174..14306 /rpt_family="MIR"
repeat_region	14635..14982 /rpt_family="MaLR"
repeat_region	14983..15321 /rpt_family="MaLR"
repeat_region	15406..15530 /rpt_family="Alu"
repeat_region	16695..16846 /rpt_family="L1"
repeat_region	16982..17188 /rpt_family="MaLR"
repeat_region	19686..19783 /rpt_family="Alu"
repeat_region	20689..21013 /rpt_family="MaLR"
repeat_region	21034..21232 /rpt_family="B4"
repeat_region	21355..21657 /rpt_family="B4"
repeat_region	21897..22051 /rpt_family="Alu"
repeat_region	22482..22598 /rpt_family="Alu"
repeat_region	22549..22616

```
repeat_region /rpt_family="B4"
23095. .23166
repeat_region /rpt_family="Alu"
23167. .23314
repeat_region /rpt_family="Alu"
23315. .23354
repeat_region /rpt_family="Alu"
23363. .23523
repeat_region /rpt_family="B4"
26232. .27488
repeat_region /rpt_family="ERVK"
27485. .27886
repeat_region /rpt_family="ERVK"
27887. .27935
repeat_region /rpt_family="B4"
28468. .28614
repeat_region /rpt_family="Alu"
28795. .28934
repeat_region /rpt_family="Alu"
30112. .30175
repeat_region /rpt_family="B2"
30193. .30526
repeat_region /rpt_family="MaLR"
30572. .30683
repeat_region /rpt_family="B2"
31165. .31186
repeat_region /rpt_family="B4"
31187. .31256
repeat_region /rpt_family="ID"
31257. .31361
repeat_region /rpt_family="B4"
34487. .34555
repeat_region /rpt_family="B4"
34561. .34705
repeat_region /rpt_family="Alu"
35095. .35294
repeat_region /rpt_family="ERVK"
36059. .36320
repeat_region /rpt_family="ERVK"
36893. .37050
repeat_region /rpt_family="B4"
37587. .37653
repeat_region /rpt_family="ERV1"
38313. .38542
repeat_region /rpt_family="B4"
38636. .38698
repeat_region /rpt_family="ERV1"
38751. .39053
repeat_region /rpt_family="MaLR"
40128. .40308
repeat_region /rpt_family="L2"
40319. .40385
repeat_region /rpt_family="B4"
40401. .40526
repeat_region /rpt_family="Alu"
40646. .40706
repeat_region /rpt_family="L2"
41898. .42130
repeat_region /rpt_family="MaLR"
42183. .42380
repeat_region /rpt_family="B2"
44039. .44238
repeat_region /rpt_family="B2"
44805. .44904
```

```
Query Match 76.8%; Score 19.2; DB 9; Length 180462;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAGGAACTCAGGCTGAG 24
||||| ||||||| |||||
Db 29952 AAAAAAATGGAACTCAGATGAG 29929
```

RESULT 47  
AC130713  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC130713 192020 bp DNA linear ROD 15-MAY-2004  
Mus musculus BAC clone RP23-316K10 from chromosome 15, complete  
sequence.  
AC130713  
AC130713.4 GI:45120394  
HTG.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
Goyea, E., Tomlinson, C., Haakenson, W. and Meyer, R.  
The sequence of Mus musculus BAC clone RP23-316K10  
Unpublished (2001)  
2 (bases 1 to 192020)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (13-AUG-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 192020)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (14-MAY-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 192020)  
Wilson, R.K.  
Direct Submission  
Submitted (05-MAR-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 192020)  
Wilson, R.K.  
Direct Submission  
Submitted (15-MAY-2004) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Mar 5, 2004 this sequence version replaced gi:30698688.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: M\_BA0316K10  
-----

NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone, fosmid clone or direct clone walk sequence.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to obtain the consensus sequence; and  
the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-23 BAC library has been constructed by Kazutoyo Osegawa  
and Minako Tatenio in the laboratory of Pieter de Jong  
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
brain genomic DNA. The clone and detailed information can be  
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
Pieter de Jong and coworkers at <http://www.chori.org>

# NEIGHBORING SEQUENCE INFORMATION: This sequence is the entire insert of the clone. This clone is overlapped by AC116950.

## FEATURES

source  
1. .192020  
/location=Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/map="15"  
/clone="RP23-316K10"  
/clone\_lib="RPCI-23"  
17082..17085  
/notes="Sequence derived from one plasmid subclone."  
misc\_feature  
17418..17861  
/notes="Sequence derived from PCR product of project DNA."  
130749..131038  
/notes="Unresolved simple sequence repeat."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 192020;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
Db 48721 AAAAACTGGAATCTCAGGCTGAG 48744

## RESULT 48

AC114552/c  
LOCUS  
DEFINITION Mus musculus chromosome 12, clone RP23-349L22, complete sequence.  
AC114552  
VERSION AC114552.16 GI:45642942  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 197585)  
Birren, B., Nusbaum, C. and Lander, E.  
Mus musculus chromosome 12, clone RP23-349L22  
Unpublished  
2 (bases 1 to 197585)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Baatien, V., Bloom, T., Bouglavkiy, L., Boukgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

JOURNAL  
REFERENCE  
3 (bases 1 to 197585)  
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
DIRECT SUBMISSION

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Baatien, V., Bloom, T., Bouglavkiy, L., Boukgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

JOURNAL  
REFERENCE  
4 (bases 1 to 197585)  
Submitted (12-FEB-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

AUTHORS  
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Baatien, V., Bloom, T., Bouglavkiy, L., Boukgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 23, 2004 this sequence version replaced gi:45339087.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L22677  
Center clone name: 349\_L\_22

## FEATURES

source  
1. .197585  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/map="12"  
/clone="RP23-349L22"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. .6845  
/notes="wgs end\_extension  
clone end:SP6"  
misc\_feature  
repeat\_region  
complement(2440..2644)

```

/rpt_family="B3"
2892..3010
/note="single clone coverage"
repeat_region
4442..4475
/rpt_family="TGGGGG)n"
repeat_region
5059..5206
/rpt_family="MIR"
complement(5391..5533)
/rpt_family="B1F"
repeat_region
6481..6511
/rpt_family="(TTGG)n"
repeat_region
complement(6524..6624)
/rpt_family="RSINE1"
6846..6851
/note="clone boundary
clone end:SP6
site:EcoRI"
repeat_region
7510..7549
/rpt_family="(TG)n"
repeat_region
7980..8352
/rpt_family="MLT1H1"
repeat_region
complement(8694..8753)
/rpt_family="L2"
repeat_region
9133..9473
/rpt_family="L1_MM"
repeat_region
9838..9993
/rpt_family="(TCTCCC)n"
repeat_region
9998..10066
/rpt_family="(TCTG)n"
complement(10161..10165)
/note="30 qual SINGL region"
11431..11597
/rpt_family="ORR1A3"
repeat_region
11900..11991
/rpt_family="MER20"
repeat_region
12010..12054
/rpt_family="(TA)n"
repeat_region
12516..12537
/rpt_family="AT_rich"
repeat_region
12576..12779
/rpt_family="Lx8"
repeat_region
13023..13082
/rpt_family="Lx8"
repeat_region
13124..13258
/rpt_family="B1_MM"
repeat_region
13259..13477
/rpt_family="(GAAA)n"
repeat_region
13484..13531
/rpt_family="(GGGA)n"
repeat_region
complement(14217..14449)
/rpt_family="Lx"
14656..14711
/rpt_family="(TC)n"
repeat_region
14711..14730
/rpt_family="(TG)n"
complement(14880..15121)
/rpt_family="ORR1D"
15621..15656
/rpt_family="(CCCCCT)n"
repeat_region
15658..15841
/rpt_family="C-rich"
17085..17182
/rpt_family="PB1D10"
repeat_region
17647..17694
/rpt_family="(CA)n"
repeat_region
17699..17905
/rpt_family="MLT1C"
complement(18654..18823)
/rpt_family="RMER10B"
18824..18902
/rpt_family="(GGAGAA)n"
repeat_region
18903..19084
/rpt_family="(GGAGAA)n"
18962..19013
/note="single clone coverage"
19014..19015
/note="unresolved VNTR: restriction enzyme digest
fingerprint indicates <100 bp of poly [GA] missing here"
complement(19165..19404)
/rpt_family="RMER10B"
repeat_region
19808..20229
/rpt_family="RMER17C"
20329..20956
/rpt_family="Lx8"
repeat_region
21021..21066
/rpt_family="(TG)n"

Query Match 76.8%; Score 19.2; DB 9; Length 197585;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
|||||
Db 49411 AAAGACTGGATCAGAGGCTGAGA 49388
|||||

RESULT 49
AC118371/c
LOCUS
DEFINITION
AC118371
AC118371.3 GI:23269853
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 197966)
Muzny, D., Marie, M., Metsker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Evans, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, S., King, L., Kovar, C.,
Kwis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensheue, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwaekeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, C.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

```

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 197966)  
Worley,K.C.

Direct Submission  
Submitted (17-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 197966)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
sequence may extend beyond the ends of the clone and there may be  
contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GTVH  
Center clone name: CH230-308G5  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 174221 bases at least Q40  
Consensus quality: 176932 bases at least Q30  
Consensus quality: 178502 bases at least Q20  
Estimated insert size: 150779; sum-of-contigs estimation  
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 50838: contig of 50838 bp in length  
\* 50839 50938: gap of unknown length  
\* 50939 57719: contig of 6781 bp in length  
\* 57720 57819: gap of unknown length  
\* 57820 196829: contig of 139010 bp in length  
\* 196830 196929: gap of unknown length  
\* 196930 197966: contig of 1037 bp in length.  
Location/Qualifiers  
1 .197966  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

FEATURES  
source

misc\_feature  
1..1556  
/clone="CH230-308G5"  
note="wgs contig"  
1607..3047  
note="wgs contig"  
50839..50938  
/estimated\_length=unknown  
57720..57819  
/estimated\_length=unknown  
57820..58944  
note="wgs contig"  
196830..196929  
/estimated\_length=unknown

gap  
misc\_feature  
50839..50938  
/estimated\_length=unknown  
57720..57819  
/estimated\_length=unknown  
57820..58944  
note="wgs contig"  
196830..196929  
/estimated\_length=unknown

ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 197966;  
Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| |||||||  
DB 31297 AAAAACTCAATCTCAGGCTGAGA 31274  
||||| |||||||

RESULT 50  
AC166021/c  
LOCUS  
DEFINITION  
Oryctolagus cuniculus clone LBI-280H21, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 19 unordered pieces.  
ACCESSION AC166021.1 GI:71037478  
VERSION AC166021.1  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Oryctolagus cuniculus (rabbit)  
ORGANISM Oryctolagus cuniculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;  
Leporidae; Oryctolagus.  
REFERENCE  
1 (bases 1 to 213357)  
Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,  
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,  
Coleman,H., Engle,J., Fuksenko,T., Gestole,M., Greene,A., Guan,X.,  
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,  
Hunter,G., Hurler,B., Idol,J.R., Kwong,P., Laric,P., Larson,S.,  
Lee-Lin,S.-O., Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B.,  
Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J., Mojidi,H.A.,  
Mullikin,J.C., Oestricher,J.S., Park,M., Portnoy,M.E., Prasad,A.,  
Puri,O., Reddix-Dugue,N., Sante,A., Schandler,K., Schueler,M.G.,  
Sison,C., Stantropop,S., Stephen,E., Taye,A., Thomas,J.W.,  
Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D.,  
Withers,T.R., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 213357)  
Green,E.D.  
Direct Submission  
Submitted (20-JUL-2005) NIH Intramural Sequencing Center, 5625  
Fishers Lane, Rockville, MD 20852, USA  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc\_zoo@nhgri.nih.gov  
----- Project Information  
Center project name: lpx  
Center clone name: 280H21  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 199123 bases at least Q40  
Consensus quality: 201873 bases at least Q30  
Consensus quality: 204164 bases at least Q20  
Insert size: 213000; agarose-fp

Insert size: 211557; sum-of-contigs  
Quality coverage: 3.94x in Q20 bases; agarose-fp  
Quality coverage: 3.96x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1  
\* 2193: contig of 2193 bp in length  
\* 2194  
\* 2293: gap of unknown length  
\* 2294  
\* 4753: contig of 2460 bp in length  
\* 4754  
\* 4853: gap of unknown length  
\* 4854  
\* 7412: contig of 2559 bp in length  
\* 7413  
\* 7512: gap of unknown length  
\* 7513  
\* 10881: contig of 3369 bp in length  
\* 10882  
\* 10881: gap of unknown length  
\* 10882  
\* 15861: contig of 4280 bp in length  
\* 15862  
\* 15361: gap of unknown length  
\* 15362  
\* 19915: contig of 4554 bp in length  
\* 19916  
\* 20015: gap of unknown length  
\* 20016  
\* 24367: contig of 4352 bp in length  
\* 24368  
\* 24467: gap of unknown length  
\* 24468  
\* 31449: contig of 6982 bp in length  
\* 31450  
\* 31549: gap of unknown length  
\* 31550  
\* 39528: contig of 7979 bp in length  
\* 39529  
\* 39628: gap of unknown length  
\* 39629  
\* 44856: contig of 5228 bp in length  
\* 44857  
\* 44956: gap of unknown length  
\* 44957  
\* 51499: contig of 6543 bp in length  
\* 51500  
\* 51599: gap of unknown length  
\* 51600  
\* 61361: contig of 9762 bp in length  
\* 61362  
\* 69454: contig of 7993 bp in length  
\* 69455  
\* 69554: gap of unknown length  
\* 69555  
\* 85346: contig of 15792 bp in length  
\* 85347  
\* 85446: gap of unknown length  
\* 85447  
\* 96665: contig of 11219 bp in length  
\* 96666  
\* 113833: contig of 17068 bp in length  
\* 113834  
\* 132574: contig of 18641 bp in length  
\* 132575  
\* 132674: gap of unknown length  
\* 132675  
\* 170486: contig of 37812 bp in length  
\* 170487  
\* 170586: gap of unknown length  
\* 170587  
\* 213357: contig of 42771 bp in length.

FEATURES  
source

1. .213357  
Location/Qualifiers  
/organism="Oryctolagus cuniculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9986"  
/clone\_lib="LB1-280H21"  
/clone\_lib="LB1"  
/note="BAC resource: <http://bacpac.chori.org/>  
Breed: New Zealand White"  
1. .2193  
/note="assembly\_fragment"  
2194. .2293  
/estimated\_length=unknown  
2294. .4753  
/note="assembly\_fragment"  
4754. .4853  
/estimated\_length=unknown  
4854. .7412  
/note="assembly\_fragment"  
7413. .7512  
/estimated\_length=unknown  
7513. .10881  
/note="assembly\_fragment"  
10882. .10981

misc\_feature

1. .2193

2194. .2293

gap

2294. .4753

misc\_feature

4754. .4853

gap

4854. .7412

misc\_feature

7413. .7512

gap

7513. .10881

misc\_feature

10882. .10981

gap

misc\_feature  
10982. .15261  
/note="assembly\_fragment"  
15262. .15361  
/estimated\_length=unknown  
15362. .19915  
/note="assembly\_fragment"  
19916. .20015  
/estimated\_length=unknown  
20016. .24367  
/note="assembly\_fragment"  
24368. .24467  
/estimated\_length=unknown  
24468. .31449  
/note="assembly\_fragment"  
31450. .31549  
/estimated\_length=unknown  
31550. .39528  
/note="assembly\_fragment"  
39529. .39628  
/estimated\_length=unknown  
39629. .44856  
/note="assembly\_fragment"  
44857. .44956  
/estimated\_length=unknown  
44957. .51499  
/note="assembly\_fragment"  
51500. .51599  
/estimated\_length=unknown  
51600. .61361  
/note="assembly\_fragment"  
61362. .61461  
/estimated\_length=unknown  
61462. .69454  
/note="assembly\_fragment"  
69455. .69554  
/estimated\_length=unknown  
69555. .85346  
/note="assembly\_fragment"  
85347. .85446  
/estimated\_length=unknown  
85447. .96665  
/note="assembly\_fragment"  
96666. .96765  
/estimated\_length=unknown  
96766. .113833  
/note="assembly\_fragment"  
113834. .113933  
/estimated\_length=unknown  
113934. .132574  
/note="assembly\_fragment"  
132575. .132674  
/estimated\_length=unknown  
132675. .170486  
/note="assembly\_fragment"  
170487. .170586  
/estimated\_length=unknown  
170587. .213357  
/note="assembly\_fragment"

ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 213357;

Best Local Similarity 87.5%; Pred. No. 2.5e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24

||||| ||||| ||||| |||||

Db 195552 AAAACCTGCAATCCAGGCTGAG 195529

Search completed: February 3, 2006, 21:27:26  
Job time : 876 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-30

Perfect score: 25

Sequence: 1 aaaaactggaatctcaggtgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : N Geneseq 21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1980s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001bs.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002bs.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	80.0	300	2	Aaz12748 Human gen
2	20	80.0	758	2	Aaz15676 Human gen
3	20	80.0	5174	6	Abn59811 Novel hum
4	20	80.0	11124	10	Adi19453 cDNA encod
5	20	80.0	11127	14	Ady17371 DNA encod
6	18.6	74.4	1484	2	Aax20314 Borrelia
7	18.6	74.4	2065	10	Ada52903 Human cod
8	18.6	74.4	2167	10	Add45029 Rat gene
9	18	72.0	956	8	Abz18441 Group III
10	17.8	71.2	876	13	Adx30549 Plant ful
11	17.8	71.2	1911	13	Ada54564 Bacteriol
12	17.8	71.2	28170	6	Abn01447 Streptoco
13	17.6	70.4	214	3	Adf57295 Urogenita
14	17.6	70.4	276	12	Adp28801 Human sec
15	17.6	70.4	522	12	Ach68960 Human gen
16	17.6	70.4	561	4	Aah09027 Human cdn
17	17.6	70.4	1266	13	Acn39840 Tumour-as
18	17.6	70.4	1688	4	Aah13739 Human cdn
19	17.6	70.4	2177	13	Acn39839 Tumour-as

c 20	17.6	70.4	3081	13	ADR07065	Adr07065 Full leng
c 21	17.6	70.4	5271	8	ABX63354	Abx63354 Human cdn
c 22	17.6	70.4	5271	10	ADJ56534	Adj56534 Rat cdna
c 23	17.6	70.4	14815	6	ABS76688	Abs76688 PRRS viru
c 24	17.6	70.4	20555	14	ADZ13631_4	Continuation (5 of
c 25	17.6	70.4	20555	14	ADZ13620_4	Continuation (5 of
c 26	17.6	70.4	24526	6	ADA43979	Ada43979 Human tra
c 27	17.6	70.4	24526	13	ADU48434	Adu48434 Human tra
c 28	17.6	70.4	26320	10	ABZ24431	Abz24431 Human con
c 29	17.6	70.4	37442	10	ADZ59528	Adz59528 Secondary
c 30	17.6	70.4	37442	14	ADZ59517	Adz59517 Secondary
c 31	17.6	70.4	37442	14	ABE96540	AbE96540 Human ILL
c 32	17.6	70.4	37996	4	ABL07876	AbL07876 Drosophill
c 33	17.6	70.4	49999	2	AAZ23896	Aaz23896 Murine LO
c 34	17.6	70.4	49999	2	AAZ23891	Aaz23891 Murine LO
c 35	17.6	70.4	63411	12	ADQ97081	AdQ97081 Mouse can
c 36	17.6	70.4	96594	10	ADC85257	AdC85257 Mouse Ptp
c 37	17.6	70.4	96595	9	ADA02777	Ada02777 Mouse Ptp
c 38	17.6	70.4	96595	10	ADB72515	AdB72515 Mouse Ptp
c 39	17.6	70.4	96595	12	ADM74372	Adm74372 Murine ca
c 40	17.6	70.4	110000	11	ACN43998_2	Continuation (3 of
c 41	17.6	70.4	110000	11	ACN45090_0	Continuation (3 of
c 42	17.6	70.4	154902	6	ABQ88198	Abq88198 Human oet
c 43	17.6	70.4	289106	14	ADZ12614	AdZ12614 Murine ca
c 44	17.6	70.4	299598	12	ADQ59380	AdQ59380 Human can
c 45	17.4	69.6	1059	13	ADO81499	AdO81499 Plant ful
c 46	17.4	69.6	2206	13	ADO83966	AdO83966 Plant ful
c 47	17.4	69.6	2800	13	ADX63913	AdX63913 Plant ful
c 48	17.4	69.6	5161	5	ABA20105	AbA20105 Human ner
c 49	17.4	69.6	35456	11	ACN43916	Acn43916 Mouse gen
c 50	17.2	68.8	460	12	ADP94079	AdP94079 Cotton ex
c 51	17.2	68.8	517	10	ADB49718	AdB49718 Primary r
c 52	17.2	68.8	517	13	ADV39311	Adv39311 Rat cardl
c 53	17.2	68.8	877	2	AAZ16499	Aaz16499 Human gen
c 54	17.2	68.8	944	13	ADR64470	Adr64470 Cotton CD
c 55	17.2	68.8	1448	13	ADR64471	Adr64471 Cotton CD
c 56	17.2	68.8	1935	4	AAK67546	Aak67546 Human imm
c 57	17.2	68.8	2372	5	ADL63428	AdL63428 Human ova
c 58	17.2	68.8	3199	13	ADR07384	Adr07384 Full leng
c 59	17.2	68.8	3670	13	ADR08219	Adr08219 Full leng
c 60	17.2	68.8	4047	12	ADN00348	Adn00348 Novel hum
c 61	17.2	68.8	4407	8	ABT23216	Abt23216 Human pro
c 62	17.2	68.8	4683	13	ADR06893	Adr06893 Full leng
c 63	17.2	68.8	5054	12	ADQ22381	AdQ22381 Human sof
c 64	17.2	68.8	6901	13	ADR84230	Adr84230 Aepergill
c 65	17.2	68.8	33486	6	ABS78976	AbS78976 E. coli C
c 66	17.2	68.8	33486	10	ADH80543	Adh80543 Escherich
c 67	17.2	68.8	35236	11	ACN44094	Acn44094 Human gen
c 68	17.2	68.8	47841	6	ABQ80553	Abq80553 Human can
c 69	17.2	68.8	76180	13	ABD33385	Abd33385 Human can
c 70	17.2	68.8	80959	8	AAZ151405	Aaz151405 Human sec
c 71	17.2	68.8	143239	12	ADQ17729	AdQ17729 Human sof
c 72	17.2	68.8	222930	6	ABR84349	AbR84349 Human cdn
c 73	17	68.0	136	4	AAZ50021	Aaz50021 Staphyloc
c 74	17	68.0	136	8	ACA17270	AcA17270 Prokaryot
c 75	17	68.0	193	4	AAI25608	Aai25608 Probe #15
c 76	17	68.0	193	4	ABA71858	AbA71858 Human foe
c 77	17	68.0	193	4	AAI52202	Aai52202 Probe #20
c 78	17	68.0	193	4	ABA37906	AbA37906 Probe #16
c 79	17	68.0	193	4	AAK46303	Aak46303 Human bon
c 80	17	68.0	193	4	AAK20234	Aak20234 Human bra
c 81	17	68.0	193	4	ABS46026	AbS46026 Human liv
c 82	17	68.0	193	6	ABS20617	AbS20617 Human liv
c 83	17	68.0	300	2	AAZ14445	Aaz14445 Human gen
c 84	17	68.0	304	3	ABC19979	AbC19979 Human sec
c 85	17	68.0	349	6	ABQ85851	Abq85851 Arabidops
c 86	17	68.0	388	6	ABN60231	Abn60231 Human can
c 87	17	68.0	397	4	AAI81444	Aai81444 Human pol
c 88	17	68.0	399	4	AAZ48855	Aaz48855 Staphyloc
c 89	17	68.0	399	8	ACA16126	AcA16126 Prokaryot
c 90	17	68.0	402	8	ABX43585	Abx43585 Bovine ES
c 91	17	68.0	416	6	ABN24330	Abn24330 Human ORF
c 92	17	68.0	456	4	AAI16428	Aai16428 Probe #63

C 93	17	68.0	456	4	ABA59346	Human foe
C 94	17	68.0	456	4	AAI39147	Probe #78
C 95	17	68.0	456	4	ABA28047	Probe #85
C 96	17	68.0	456	4	AAK33360	Human bon
C 97	17	68.0	456	4	AAK07564	Human bra
C 98	17	68.0	456	4	ABg33114	Human liv
C 99	17	68.0	456	6	ABs08195	Human gen
C 100	17	68.0	464	5	ABv09920	Human pro
C 101	17	68.0	471	2	AAe68141	H. pylori
C 102	17	68.0	477	12	ADL84731	DNA up-re
C 103	17	68.0	477	12	ADL84732	DNA up-re
C 104	17	68.0	486	9	ACH23112	Human adu
C 105	17	68.0	489	4	AAe51634	Staphyloc
C 106	17	68.0	498	8	ACF74024	Staphyloc
C 107	17	68.0	501	4	AAe54715	Staphyloc
C 108	17	68.0	501	8	ACA19754	Prokaryot
C 109	17	68.0	501	14	ADW94293	Prolifera
C 110	17	68.0	503	4	AAK64418	Human imm
C 111	17	68.0	522	10	ADB37434	Human can
C 112	17	68.0	561	13	AQO50681	Novel can
C 113	17	68.0	631	3	AAc10899	Human sec
C 114	17	68.0	807	10	ACD19376	CDNA enco
C 115	17	68.0	858	2	AAV25026	H. pylori
C 116	17	68.0	864	2	AAV24698	H. pylori
C 117	17	68.0	890	14	ADx25862	Novel cel
C 118	17	68.0	911	6	ABN87869	Human ova
C 119	17	68.0	968	13	ADx31894	Plant ful
C 120	17	68.0	970	6	ABN87870	Human ova
C 121	17	68.0	971	13	ADx12208	Plant ful
C 122	17	68.0	1002	4	AAK70298	Human imm
C 123	17	68.0	1004	10	ADB58326	Toxicity-
C 124	17	68.0	1004	10	ADB52885	Primary r
C 125	17	68.0	1004	10	ABT42005	Toxicity
C 126	17	68.0	1004	13	ADV41152	Rat cardi
C 127	17	68.0	1004	14	ADx26224	Novel cel
C 128	17	68.0	1043	10	ADB37435	Human can
C 129	17	68.0	1074	14	ADM16706	Eucalyptu
C 130	17	68.0	1105	13	ACN39822	Tumour-as
C 131	17	68.0	1108	2	AAx13460	Enterococ
C 132	17	68.0	1108	6	ABs99255	Enterococ
C 133	17	68.0	1111	5	AAH76472	CDNA corr
C 134	17	68.0	1111	5	AAH76472	CDNA corr
C 135	17	68.0	1111	6	ABL61814	Colon ade
C 136	17	68.0	1111	6	ABF92125	Human tum
C 137	17	68.0	1111	8	ACF12940	Human cer
C 138	17	68.0	1111	10	ACD19378	CDNA enco
C 139	17	68.0	1111	10	ABx72050	DNA enco
C 140	17	68.0	1138	10	ADK65794	Angiogene
C 141	17	68.0	1159	12	ADJ76032	Marker ge
C 142	17	68.0	1220	3	AAc45643	Arabidops
C 143	17	68.0	1222	3	AAc47777	Arabidops
C 144	17	68.0	1265	10	ADe15673	Human str
C 145	17	68.0	1299	8	ADA70923	Rice gene
C 146	17	68.0	1424	2	AAV74877	Staphyloc
C 147	17	68.0	1484	10	ADe15659	Human str
C 148	17	68.0	1512	13	ADx52319	Plant ful
C 149	17	68.0	1538	13	ADx48874	Plant ful
C 150	17	68.0	1593	13	ADs00107	TM5a iso

ALIGNMENTS

RESULT 1  
AAZ12748  
ID AAZ12748 standard; cDNA; 300 BP.

AAZ12748;

12-OCT-1999 (first entry)

Human gene expression product cDNA sequence SEQ ID NO:217.

KW	Human; gene; gene expression product; diagnosis; therapy; probe; detection; mapping; tissue typing; profiling; forensic; cancer; genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
KW	Homo sapiens.
XX	WO9938972-A2.
XX	05-AUG-1999.
XX	28-JAN-1999; 99US-0001619.
XX	28-JAN-1998; 98US-0072910P.
XX	24-FEB-1998; 98US-0075954P.
XX	31-MAR-1998; 98US-0080114P.
XX	03-APR-1998; 98US-0080515P.
XX	03-APR-1998; 98US-0080666P.
XX	21-OCT-1998; 98US-0105234P.
XX	28-OCT-1998; 98US-0105877P.
XX	(CHIR ) CHIRON CORP.
XX	(HYSE-) HYSEQ INC.
XX	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J; Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A; Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I; Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B; WPI; 1999-494092/41.
XX	Novel human genes and their expression products which are differentially expressed in different cell types.
XX	Claim 1; Page 712; 2479pp; English.
XX	The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ17779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purposes, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists
XX	Sequence 300 BP; 99 A; 58 C; 49 G; 94 T; 0 U; 0 Other;
XX	Query Match 80.0%; Score 20; DB 2; Length 300;
XX	Best Local Similarity 100.0%; Pred. No. 34;
XX	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	4 AAACCTGGAAATCTCAGGCTGA 23
Db	104 AAACCTGGAAATCTCAGGCTGA 123
RESULT 2	
AAZ15676	
ID AAZ15676 standard; cDNA; 758 BP.	
XX	
AC	AAZ15676;

```
XX 12-OCT-1999 (first entry)
DT Human gene expression product cDNA sequence SEQ ID NO:3145.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
OS
XX WO9938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US001619.
XX
XX 28-JAN-1998; 98US-0072910P.
XX
XX 24-FEB-1998; 98US-0075954P.
XX
XX 31-MAR-1998; 98US-0080114P.
XX
XX 03-APR-1998; 98US-0080515P.
XX
XX 03-APR-1998; 98US-0080666P.
XX
XX 21-OCT-1998; 98US-0105234P.
XX
XX 28-OCT-1998; 98US-0105877P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are differentially
PT expressed in different cell types.
XX
XX Claim 1; Page 1511; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one of
CC the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists
XX
XX Sequence 758 BP; 221 A; 146 C; 148 G; 215 T; 0 U; 28 Other;
SQ
Query Match 80.0%; Score 20; DB 2; Length 758;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 AAACCTGGAATCTCAGGCTGA 23
Db 148 AAACCTGGAATCTCAGGCTGA 167
```

```
RESULT 3
ABN59811
ID ABN59811 standard; cDNA; 5174 BP.
XX
XX AC ABN59811;
XX
XX 28-JUN-2002 (first entry)
DT
XX
XX Novel human coding sequence SEQ ID NO: 222.
DE
XX
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX
XX Homo sapiens.
OS
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US026015.
XX
XX 11-SEP-2000; 2000US-00659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI; 2002-292408/33.
DR
XX P-PSDB; ABB97398.
XX
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
XX Claim 1; SEQ ID NO 222; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g. of the
CC Parkinson's disease. The present sequence is a coding sequence of the
XX invention
XX
XX Sequence 5174 BP; 1361 A; 1184 C; 1262 G; 1367 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 20; DB 6; Length 5174;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 AAACCTGGAATCTCAGGCTGA 23
Db 964 AAACCTGGAATCTCAGGCTGA 983
RESULT 4
ACD19453
ID ACD19453 standard; cDNA; 11124 BP.
XX
XX AC ACD19453;
XX
XX 25-AUG-2003 (first entry)
DT
XX
XX cDNA encoding novel human protein #133.
DE
XX Human; NOV; gene therapy; endocrine related disease; diabetes;
KW metabolism-related disease; obesity; central nervous system disorder;
```

KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;  
KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;  
KW psoriasis; allergy; lupus erythematosus; asthma; cancer;  
KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;  
KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;  
KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;  
KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;  
KW stroke; infection; gene; ss.

OS Homo sapiens.

XX  
XX WO2003023002-A2.

XX  
XX 20-MAR-2003.

XX  
XX 09-SEP-2002; 2002WO-US028539.

XX  
XX 07-SEP-2001; 2001US-0318120P.

XX  
XX 07-SEP-2001; 2001US-0318130P.

XX  
XX 10-SEP-2001; 2001US-0318430P.

XX  
XX 17-SEP-2001; 2001US-0322836P.

XX  
XX 17-SEP-2001; 2001US-0322781P.

XX  
XX 17-SEP-2001; 2001US-0322816P.

XX  
XX 17-SEP-2001; 2001US-0322817P.

XX  
XX 19-SEP-2001; 2001US-0323519P.

XX  
XX 20-SEP-2001; 2001US-0323519P.

XX  
XX 20-SEP-2001; 2001US-0323631P.

XX  
XX 20-SEP-2001; 2001US-0323636P.

XX  
XX 25-SEP-2001; 2001US-0324969P.

XX  
XX 25-SEP-2001; 2001US-0325091P.

XX  
XX 26-SEP-2001; 2001US-0324990P.

XX  
XX 17-APR-2002; 2002US-0173212P.

XX  
XX 06-SEP-2002; 2002US-00236177.

XX  
XX (CURA-) CURAGEN CORP.

XX  
XX Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;

XX  
XX Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;

XX  
XX Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;

XX  
XX Riegler DK, Taupier RJ, Shenoy SG, Liu X, Padigar M, Alsobrook JP;

XX  
XX Lepley DM, Edinger SR, Burgess CE;

XX  
XX WPI; 2003-313242/30.

XX  
XX P-PSDB; ABO14760.

XX  
XX New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)

XX  
XX and polynucleotides, useful in gene therapy, e.g. for treating or

XX  
XX preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,

XX  
XX stroke or infections.

XX  
XX Claim 20; Page 373-375; 586pp; English.

XX  
XX The invention describes a new isolated polypeptide (NOVX). The NOVX

XX  
XX polypeptide, nucleic acid and antibody are useful as therapeutics,

XX  
XX particularly in the manufacture of a medicament for treating a syndrome

XX  
XX associated with a human disease, which includes a pathology associated

XX  
XX with NOVX polypeptide. The DNA encoding the protein is useful in gene

XX  
XX therapy for treating the disease or condition. In particular, the NOVX

XX  
XX polypeptide or polynucleotide is useful for treating endocrine/

XX  
XX metabolism-related diseases (e.g. obesity or diabetes), central nervous

XX  
XX system disorders (e.g. Alzheimer's disease, Parkinson's disease,

XX  
XX epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune

XX  
XX and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,

XX  
XX asthma, inflammatory bowel disease, rheumatoid arthritis or

XX  
XX osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,

XX  
XX prostate or brain cancers, or melanoma), liver diseases (e.g. liver

XX  
XX cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),

XX  
XX haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).

XX  
XX These are also useful in developing powerful assay system for functional

XX  
XX analysis of various human disorders, as well as in diagnostic

XX  
XX applications, and for monitoring the effects of drugs during clinical

XX  
XX trials. This sequence encodes a novel human NOV protein

XX  
XX Sequence 11124 BP; 2855 A; 2715 C; 2624 G; 2928 T; 0 U; 2 Other;

XX  
XX

XX  
XX

XX  
XX

XX  
XX

XX  
XX

XX  
XX

XX  
XX

XX  
XX

Query Match 80.0%; Score 20; DB 10; Length 11124;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 AAACCTGGAATCTCAGGCTGA 23  
|||||  
DB 964 AAACCTGGAATCTCAGGCTGA 983

RESULT 5

ADY17371

ID ADY17371 standard; DNA; 11127 BP.

XX

AC ADY17371;

XX

DT 05-MAY-2005 (first entry)

XX

XX DNA encoding a PRO polypeptide, SEQ ID NO 3177.

XX

XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;

XX Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;

XX Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;

XX Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;

XX Antiallergic; ds; gene; diagnosis.

XX

OS Homo sapiens.

XX

XX WO2005016962-A2.

XX

XX 24-FEB-2005.

XX

XX 11-AUG-2004; 2004WO-US026249.

XX

XX 11-AUG-2003; 2003US-0493546P.

XX (GETH ) GENENTECH INC.

XX PA Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;

XX PI

XX WPI; 2005-182330/19.

XX

XX New nucleic acid encoding PRO polypeptide, useful for diagnosing and

XX PT treating an immune related disorder, e.g. systemic lupus erythematosus,

XX PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.

XX

XX Claim 1; SEQ ID NO 3177; 158pp; English.

XX

XX The invention relates to an isolated nucleic acid encoding a PRO

XX polypeptide. The polypeptide, agonist or an antagonist, antibody,

XX composition, and method are useful for diagnosing and treating an immune

XX related disorder, e.g. systemic lupus erythematosus, rheumatoid

XX arthritis. The present sequence represents a DNA encoding a PRO

XX polypeptide.

XX

SQ Sequence 11127 BP; 2858 A; 2712 C; 2628 G; 2929 T; 0 U; 0 Other;

XX

XX Query Match 80.0%; Score 20; DB 14; Length 11127;

XX Best Local Similarity 100.0%; Pred. No. 60;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

QY 4 AAACCTGGAATCTCAGGCTGA 23

|||||

DB 964 AAACCTGGAATCTCAGGCTGA 983

XX

RESULT 6

AA20314/c

ID AAX20314 standard; DNA; 1484 BP.

XX

XX AAX20314;

XX

DT 04-MAY-1999 (first entry)

XX

```

XX  Borrelia burgdorferi polynucleotide sequence #67.
DE
XX
XX  Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW  epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW  infection; diagnosis; characterisation; detection; ds.
XX
XX  Borrelia burgdorferi.
OS
XX
XX  WO9858943-A1.
PN
XX
XX  30-DEC-1998.
PD
XX
XX  18-JUN-1998; 98WO-US012764.
PF
XX
XX  20-JUN-1997; 97US-0050359P.
XX
XX  22-JUL-1997; 97US-0053344P.
PR
XX  22-JUL-1997; 97US-0053377P.
PR
XX  03-SEP-1997; 97US-0057483P.
PR
XX
XX  (HUMA-) HUMAN GENOME SCI INC.
XX
XX  (MEDI-) MEDIMUNE INC.
PA
XX
XX  Fraser C, White OR, Clayton R, Dougherty BA, Lathigra R;
PI  Smith HO;
PI
XX
XX  WPI; 1999-081217/07.
XX
XX  New isolated Borrelia burgdorferi nucleic acids - used to develop
PT  products for the detection, diagnosis, characterisation, prevention and
PT  therapy of infections, particularly Lyme disease.
XX
XX  Claim 1; Page 1050-1051; 1128pp; English.
XX
XX  AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC  Borrelia burgdorferi (Bb). Products derived from Bb can be used for the
CC  detection, diagnosis, characterisation, prevention and therapy of Bb
CC  infections, e.g. Lyme disease. They can also be used for the production
CC  of biosynthetic products, e.g. enzymes. Borrelia belongs to a family of
CC  motile, spiral-shaped bacteria called Spirochetes. Spirochetes are
CC  pathogenic in humans and Borrelia causes epidemic and endemic relapsing
CC  fever, and Lyme borreliosis, more commonly known as Lyme disease
XX
XX  Sequence 1484 BP; 410 A; 290 C; 208 G; 571 T; 0 U; 5 Other;
SQ
XX
Query Match 74.4%; Score 18.6; DB 2; Length 1484;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX
XX  1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 454 AAAAACTGCAAGAACTCGGCTGAGA 430

RESULT 7
ADA52903
ID ADA52903 standard; cDNA; 2066 BP.
XX
XX  ADA52903;
AC
XX
XX  20-NOV-2003 (first entry)
DT
XX
XX  Human coding sequence, SEQ ID 471.
DE
XX
XX  Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW  Gene therapy; human; secretory protein; membrane proteins; cancer;
KW  inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
XX  Homo sapiens.
OS
XX
XX  EP1293569-A2.
PN
XX
XX  19-MAR-2003.
PD

```

```

XX  21-MAR-2002; 2002EP-00006586.
PF
XX
XX  14-SEP-2001; 2001JP-00328381.
PR
XX  24-JAN-2002; 2002US-0350435P.
PR
XX
XX  (HELI-) HELIX RES INST.
PA
XX  (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX  Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI  Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI  Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX  WPI; 2003-395539/38.
DR
XX  P-PSDB; ADA54542.
DR
XX
XX  New polynucleotides encoding full-length polypeptides, e.g. secretory
PT  and/or membrane proteins, useful for developing medicines for diseases in
PT  which the gene is involved, or as target molecules for gene therapy.
XX
XX  Claim 1; SEQ ID NO 471; 205pp; English.
XX
XX  The present invention relates to novel human secretory or membrane
CC  proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC  ADA54071). The coding sequences are useful in the gene therapy of
CC  diseases caused by abnormalities of the proteins, e.g. cancer,
CC  inflammatory diseases, osteoporosis or neurological disease.
XX
XX  Sequence 2066 BP; 460 A; 696 C; 537 G; 373 T; 0 U; 0 Other;
SQ
XX
Query Match 74.4%; Score 18.6; DB 10; Length 2066;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX
XX  1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 820 AAAAACTCGAAACTCGGCTGAGA 844

RESULT 8
ADA5029/c
ID ADA5029 standard; DNA; 2167 BP.
XX
XX  ADA5029;
AC
XX
XX  29-JAN-2004 (first entry)
DT
XX
XX  Rat gene Z48444, SEQ ID NO 10461.
DE
XX
XX  Rat; ds; Gene; pain; neuronal tissue; gene therapy;
KW  spinal segmental nerve injury; chronic constriction injury; CCI;
KW  spared nerve injury; SNI; Chung.
XX
XX  Rattus norvegicus.
OS
XX
XX  WO2003016475-A2.
PN
XX
XX  27-FEB-2003.
PD
XX
XX  14-AUG-2002; 2002WO-US025765.
XX
XX  14-AUG-2001; 2001US-0312147P.
PR
XX  01-NOV-2001; 2001US-0346382P.
PR
XX  26-NOV-2001; 2001US-0333347P.
PR
XX
XX  (GEHO ) GEN HOSPITAL CORP.
PA
XX  (FARB ) BAYER AG.
XX
XX  Woolf C, D'urso D, Befort K, Costigan M;
PI
XX
XX  WPI; 2003-268312/26.
DR
XX  GENBANK; Z48444.
XX

```

PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2167 BP; 611 A; 441 C; 510 G; 605 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 10; Length 2167;  
Best Local Similarity 84.0%; Pred.No. 2.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGATCTCAGGCTGAGA 25

Db 1642 AAAAAATTGGAATTCAGGCTTAAA 1618

RESULT 9

ABZ18441

ID ABZ18441 standard; cDNA; 956 BP.

AC ABZ18441;

XX 23-JAN-2003 (first entry)

XX Group III CDNA cancer related clone SEQ ID NO:867.

XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;  
KW immune response; virology; immunology; microbiology; molecular biology;  
KW recombinant DNA technology; gene; ss.

XX Homo sapiens.

XX WO200278516-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US010421.

XX 30-MAR-2001; 2001US-0280255P.

XX 28-AUG-2001; 2001US-0315563P.

XX 09-JAN-2002; 2002US-0347313P.

XX (CORI-) CORIXA CORP.

PI Wang T, Wang S, Bangur CS, Gaiger A;  
XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,  
PT preventing and treating cancer expressing CT or CP mRNA antigens, and in  
PT virology, immunology, microbiology, molecular biology and recombinant DNA  
PT techniques.

PS Claim 1; SEQ ID NO 867; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and  
CC ABP54446 to ABP54472 represent protein (II) sequences, from the present  
CC invention. (I) and (II) have cytostatic activity and can be used in gene  
CC therapy and vaccines. (I), (II), antibodies and compositions from the  
CC present invention are useful for diagnosing, preventing and treating  
CC cancer, which expresses CT or CP mRNA antigens. They are useful for  
CC stimulating immune response. They can also be useful in virology,  
CC immunology, microbiology, molecular biology and recombinant DNA  
CC techniques. N.B. The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 956 BP; 273 A; 168 C; 189 G; 301 T; 0 U; 25 Other;

Query Match 72.0%; Score 18; DB 8; Length 956;  
Best Local Similarity 100.0%; Pred.No. 3.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGC 20

Db 687 AAAAAGTGAATCTCAGGC 704

RESULT 10

ADX30549/c

ID ADX30549 standard; cDNA; 876 BP.

XX AC ADX30549;

XX 21-APR-2005 (first entry)

XX Plant full length insert polynucleotide seqid 13369.

XX plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.

XX Unidentified.

XX US2004034888-A1.

XX 19-FEB-2004.

XX 28-APR-2003; 2003US-00425114.

XX 06-MAY-1999; 99US-00304517.

XX 05-NOV-2001; 2001US-00985678.

XX (LIUJ/) LIU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABA/) TABASKA J E.

XX (CAOY/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

DR WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
XX Claim 1; SEQ ID NO 13369; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 876 BP; 257 A; 145 C; 193 G; 281 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 13; Length 876;  
Best Local Similarity 90.5%; Pred. No. 4.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 4 AACTGGAATCTCAGGCTGAG 24  
Db 258 AACTGCAGTCTCAGGCTGAG 238  
||||| |||||||  
RESULT 11  
ADS45644  
ID ADS45644 standard; cDNA; 1911 BP.  
XX  
AC ADS45644;  
XX  
XX 02-DEC-2004 (first entry)  
XX Bacterial polynucleotide #387.  
XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.  
XX  
XX Bacteria.  
XX  
XX US2003233675-A1.  
XX  
XX 18-DEC-2003.  
XX  
XX 20-FEB-2003; 2003US-00369493.  
XX  
XX 21-FEB-2002; 2002US-0360039P.  
XX  
XX (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
XX WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 24074; 122pp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at segdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 1911 BP; 575 A; 431 C; 484 G; 421 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 13; Length 1911;  
Best Local Similarity 90.5%; Pred. No. 4.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 AAAACTGGAATCTCAGGCTGA 23  
Db 909 AAAACTGGAATCTCGTTGA 929  
||||| |||||||  
RESULT 12  
ABA01447  
ID ABA01447 standard; DNA; 28170 BP.  
XX  
XX ABA01447;  
AC  
XX 21-FEB-2002 (first entry)  
DT  
XX Streptococcus thermophilus eps10 operon #2.  
DE  
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
KW ds.  
XX  
XX Streptococcus thermophilus.  
OS  
XX WO200179500-A2.  
PN  
XX 25-OCT-2001.  
PD  
XX 18-APR-2001; 2001WO-FR001199.  
PF  
XX 18-APR-2000; 2000FR-00004972.  
PR  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
PA (DANO-) CIE DANONE SA GERVAIS.  
PA (RHOD ) RHODIA CHIM.  
XX



PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;  
XX WPI; 2002-017616/02.  
XX  
XX New nucleic acid fragments containing exopolysaccharide operon, useful  
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.  
PT  
XX  
XX Claim 9; Page 109-117; 144pp; French.  
XX  
XX The present sequence is an eps operon from *Streptococcus thermophilus*.  
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)  
CC synthesis. The operon is useful for producing chimeric eps operons, for  
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,  
CC mouth feel and rheological properties to fermented food products (e.g.  
CC yoghurt). They function as thickeners, to provide free-flowing and creamy  
CC texture, and may also have biological activities beneficial to health  
XX  
SQ Sequence 28170 BP; 8951 A; 4464 C; 5468 G; 9287 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 6; Length 28170;  
Best Local Similarity 90.5%; Pred. No. 7.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AAAAAGCTGGAATCTCAGGCTG 22  
DB 7501 AAAAAGCTGGAATCTCAGGCTG 7521  
RESULT 13  
ADF57295/C  
ID ADF57295 standard; cDNA; 214 BP.  
XX  
XX ADF57295;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Urogenital sinus-derived expressed sequence tag, SEQ ID NO 614.  
XX  
XX foetal urogenital sinus-derived expressed sequence tag; EST; g092; ug093;  
KW ug101; ug102; ug106; ug120; ug254; ug291; ug307; ug308; ug311; ug317;  
KW ug320; ug334; ug335; ug353; ug354; ug357; ug440; ug441; ug482; ug484;  
KW ug485; ug491; ug493; ug494; ug503; ug505; ug506; ug518; ug5186; ug5194;  
KW cytostatic; anti-inflammatory; prostate disease; prostatitis; benign;  
KW malignant; prostate gland; ss; mouse; murine.  
XX  
XX Mus sp.  
XX  
XX WO9958665-A2.  
XX  
XX 18-NOV-1999.  
XX  
XX 14-MAY-1999; 99WO-US010746.  
XX  
XX 14-MAY-1998; 98US-0085383P.  
XX  
XX (UYVI-) UNIV VIRGINIA PATENT FOUND.  
XX  
XX Silkes RA, Chung LWK, Kim JH, Fasciana C, Trapman J;  
PI WPI; 2000-039105/03.  
XX  
XX New urogenital sinus-derived expressed sequence tags, used to, e.g. treat  
PT prostate cancer.  
PT  
XX  
XX Disclosure; SEQ ID NO 614; 605pp; English.  
XX  
XX The invention relates to novel foetal urogenital sinus-derived expressed  
CC sequence tags. The invention more specifically comprises an isolated  
CC polynucleotide comprising a nucleotide sequence containing a urogenital  
CC sinus-derived expressed sequence tag (EST) comprising g092, ug093, ug101,  
CC ug102, ug106, ug120, ug254, ug291, ug307, ug308, ug311, ug317, ug320,  
CC ug334, ug335, ug353, ug354, ug357, ug440, ug441, ug482, ug484, ug485,  
CC ug491, ug493, ug494, ug503, ug505, ug506, ug518, ug5186 and ug5194. The

CC urogenital sinus-derived expressed sequence tags have cytostatic and anti  
CC -inflammatory activities. The genes containing urogenital sinus-derived  
CC ESTs encode oncofetal proteins that are useful for diagnosis of prostate  
CC disease and for monitoring the efficacy of a compound in clinical trials  
CC for the treatment of prostate disease. The ESTs can be used in assays to  
CC identify substances (e.g. triple helix forming oligonucleotides,  
CC antisense oligonucleotides, ribozymes or antibodies) useful for treating  
CC prostate disease, such as prostatitis and benign or malignant growth of  
CC the prostate gland. This polynucleotide sequence represents the cDNA of a  
CC urogenital sinus-derived expressed sequence tag of the invention.  
XX  
SQ Sequence 214 BP; 36 A; 68 C; 40 G; 70 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 3; Length 214;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAAGCTGGAATCTCAGGCTGAGA 25  
DB 58 AATCACTGGAGGCTCAGGCTGAGA 35  
RESULT 14  
ADP28801/C  
ID ADP28801 standard; DNA; 276 BP.  
XX  
XX ADP28801;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE Human secreted protein encoding sequence SEQ ID #799.  
XX  
XX Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;  
KW cancer; inflammatory; immune; ds; human secreted protein.  
XX  
XX Homo sapiens.  
XX  
XX WO2004035732-A2.  
XX  
XX 29-APR-2004.  
XX  
XX 28-AUG-2003; 2003WO-US026780.  
XX  
XX 29-AUG-2002; 2002US-0406576P.  
XX  
XX 29-AUG-2002; 2002US-0406579P.  
XX  
XX 29-AUG-2002; 2002US-0406585P.  
XX  
XX 29-AUG-2002; 2002US-0406588P.  
XX  
XX 29-AUG-2002; 2002US-0406609P.  
XX  
XX 29-AUG-2002; 2002US-0406611P.  
XX  
XX 29-AUG-2002; 2002US-0406612P.  
XX  
XX 29-AUG-2002; 2002US-0406616P.  
XX  
XX 29-AUG-2002; 2002US-0406640P.  
XX  
XX 29-AUG-2002; 2002US-0406642P.  
XX  
XX 29-AUG-2002; 2002US-0406646P.  
XX  
XX 29-AUG-2002; 2002US-0406653P.  
XX  
XX 29-AUG-2002; 2002US-0406655P.  
XX  
XX 29-AUG-2002; 2002US-0406666P.  
XX  
XX 17-SEP-2002; 2002US-0410946P.  
XX  
XX 17-SEP-2002; 2002US-0410947P.  
XX  
XX 17-SEP-2002; 2002US-0410949P.  
XX  
XX 17-SEP-2002; 2002US-0410953P.  
XX  
XX 17-SEP-2002; 2002US-0410957P.  
XX  
XX 17-SEP-2002; 2002US-0410958P.  
XX  
XX 17-SEP-2002; 2002US-0410959P.  
XX  
XX 17-SEP-2002; 2002US-0410960P.  
XX  
XX 17-SEP-2002; 2002US-0410961P.  
XX  
XX 17-SEP-2002; 2002US-0410962P.  
XX  
XX 17-SEP-2002; 2002US-0411019P.  
XX  
XX 17-SEP-2002; 2002US-0411022P.  
XX  
XX 17-SEP-2002; 2002US-0411023P.  
XX  
XX 17-SEP-2002; 2002US-0411024P.  
XX  
XX 17-SEP-2002; 2002US-0411032P.

PR 17-SEP-2002; 2002US-0411035P.  
PR 17-SEP-2002; 2002US-0411037P.  
PR 17-SEP-2002; 2002US-0411041P.  
PR 17-SEP-2002; 2002US-0411041P.  
PR 17-SEP-2002; 2002US-0411045P.  
PR 17-SEP-2002; 2002US-0411046P.  
PR 17-SEP-2002; 2002US-0411048P.  
PR 17-SEP-2002; 2002US-0411052P.  
PR 17-SEP-2002; 2002US-0411055P.  
PR 17-SEP-2002; 2002US-0411073P.  
PR 17-SEP-2002; 2002US-0411082P.  
PR 17-SEP-2002; 2002US-0411101P.  
PR 17-SEP-2002; 2002US-0411111P.  
PR 18-APR-2003; 2003US-0463700P.  
PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463716P.  
PR 18-APR-2003; 2003US-0463732P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467201P.  
PR 02-MAY-2003; 2003US-0467203P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 22-MAY-2003; 2003US-0472420P.  
PR 22-MAY-2003; 2003US-0472430P.  
PR 03-JUN-2003; 2003US-0476609P.  
PR 03-JUN-2003; 2003US-0476641P.  
PR 08-JUL-2003; 2003US-0485218P.  
PR 08-JUL-2003; 2003US-0485223P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 08-JUL-2003; 2003US-0485325P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 14-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 15-JUL-2003; 2003US-0486960P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.  
PR 08-AUG-2003; 2003US-0493573P.  
PR 08-AUG-2003; 2003US-0493577P.  
XX  
XX  
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX  
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;  
PI Halenbeck RE, Huang MM, Kothakota S, Haishan L, Linnemann T;  
PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;  
XX  
XX WPI; 2004-348438/32.  
XX  
XX New nucleic acid molecule for diagnosing, preventing or treating diseases  
PT such as proliferative (e.g. cancer), inflammatory, immune, metabolic,  
PT genetic, bacterial and viral diseases.  
XX  
XX Claim 1; SEQ ID NO 799; 428pp; English.  
XX  
XX The present invention relates to an isolated nucleic acid molecule  
CC encoding a polypeptide which is believed to be cytostatic,  
CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The  
CC composition and methods are useful for diagnosing, preventing and  
CC treating diseases such as proliferative (e.g. cancer), inflammatory,  
CC immune, metabolic, genetic, bacterial and viral diseases. The present  
CC sequence represents a human secreted protein encoding sequence. The  
CC present sequence is available on WIPOMEB and is not in the specification.  
XX  
XX Sequence 276 BP; 95 A; 57 C; 45 G; 79 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 12; Length 276;  
Best Local Similarity 83.3%; Pred. No. 4.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 AAAAACTGGAATCTCAGGCTGAG 24  
|||||  
Db 232 AAAAACTGGAATCTCAGGCTGAG 209  
|||||

RESULT 15  
ACH68960/c  
ID ACH68960 standard; DNA; 522 BP.  
XX  
AC ACH68960;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human genome derived single exon probe #2155.  
XX  
KW Human; probe; ss; gene expression; single exon probe; microarray;  
KW alternative splicing event; genomic alteration.  
XX  
OS Homo sapiens.  
XX  
PN US2003194704-A1.  
XX  
PD 16-OCT-2003.  
XX  
XX 03-APR-2002; 2002US-00029386.  
XX  
XX 03-APR-2002; 2002US-00029386.  
XX  
XX (PENN/) PENN S G.  
XX (RANK/) RANK D R.  
XX (HANZ/) HANZEL D K.  
XX  
XX Penn SG, Rank DR, Hanzel DK;  
XX  
XX WPI; 2004-119264/12.  
XX  
XX New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
XX Claim 15; SEQ ID NO 2155; 80pp; English.  
XX  
XX The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subscription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterising  
CC alternative splicing events, in detecting and characterising gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704

XX SQ Sequence 522 BP; 153 A; 126 C; 68 G; 175 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 12; Length 522;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
Db 264 AAAAACTGGAATGTAAAGCTGTG 241  
RESULT 16  
AAH09027  
ID AAH09027 standard; cDNA; 561 BP.  
XX AC AAH09027;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA clone (3'-primer) SEQ ID NO:5862.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-00116126.  
XX PR 29-JUL-1999; 99JP-00248036.  
XX PR 27-AUG-1999; 99JP-00300253.  
XX PR 11-JAN-2000; 2000JP-00118776.  
XX PR 02-MAY-2000; 2000JP-00183767.  
XX PR 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 3; SEQ ID NO 5862; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the

CC present invention  
XX SQ Sequence 561 BP; 163 A; 134 C; 121 G; 137 T; 0 U; 6 Other;  
Query Match 70.4%; Score 17.6; DB 4; Length 561;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
Db 247 AAAGCACTGGAATTACAGGCTGAG 270  
RESULT 17  
ACN39840  
ID ACN39840 standard; cDNA; 1266 BP.  
XX AC ACN39840;  
XX DT 18-NOV-2004 (first entry)  
XX DE Tumour-associated antigenic target (TAT) cDNA DNA326021, SEQ ID NO:4237.  
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer; tumour; diagnosis; cell proliferative disorder; breast cancer; colorectal cancer; lung cancer; ovarian cancer; liver cancer; central nervous system cancer; bladder cancer; pancreatic cancer; cervical cancer; melanoma; leukaemia; hybridisation probe; chromosome identification; chromosome mapping; gene mapping; gene therapy; cytostatic; gene; ss.  
XX OS Homo sapiens.  
XX PN WO2004030615-A2.  
XX PD 15-APR-2004.  
XX PF 29-SEP-2003; 2003WO-US028547.  
XX PR 02-OCT-2002; 2002US-0414971P.  
XX PA (GETH ) GENENTECH INC.  
XX PI Wu TD, Zhang Z, Zhou Y;  
XX WPI; 2004-347921/32.  
XX DR P-PSDB; ABM81641.  
XX PT New tumor-associated antigenic target polypeptides and nucleic acids, useful in preparing a medicament for treating or detecting a proliferative disorder, e.g. breast, lung, colorectal, ovarian or prostate cancer or tumor.  
XX  
PS Claim 1; SEQ ID NO 4237; 7273pp; English.  
XX  
CC The invention relates to human tumour-associated antigenic target (TAT) polypeptides, and their related nucleic acids. The TAT polypeptides are overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in

CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT nucleic acid of the invention  
 XX  
 SQ Sequence 1266 BP; 432 A; 210 C; 302 G; 322 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 13; Length 1266;  
 Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
 |||||  
 DB 883 AAAAACTGGAATGTAAAGCTGTG 906  
 |||||  
 RESULT 18  
 AAH13739/C  
 ID AAH13739 standard; cDNA; 1688 BP.  
 XX  
 AC AAH13739;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:10641.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 03-JUN-2000; 2000JP-00241899.  
 XX  
 (HELI-) HELIX RES INST.  
 XX  
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 Claim 8; SEQ ID NO 10641; 2537pp + Sequence Listing; English.  
 XX  
 The present invention describes primer sets for synthesising 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 1688 BP; 353 A; 507 C; 366 G; 462 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 4; Length 1688;  
 Best Local Similarity 83.3%; Pred. No. 5.8e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
 |||||  
 DB 1442 AAAGCACTGGAATTACAGGCTGAG 1419  
 |||||  
 RESULT 19  
 ACN39839  
 ID ACN39839 standard; cDNA; 2177 BP.  
 XX  
 AC ACN39839;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Tumour-associated antigenic target (TAT) cDNA DNA326020, SEQ ID NO:4236.  
 XX  
 KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KW central nervous system cancer; bladder cancer; pancreatic cancer;  
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
 KW chromosome identification; chromosome mapping; gene mapping;  
 KW gene therapy; cytostatic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004030615-A2.  
 XX  
 PD 15-APR-2004.  
 XX  
 PF 29-SEP-2003; 2003WO-US028547.  
 XX  
 PR 02-OCT-2002; 2002US-0414971P.  
 XX  
 (GETH ) GENENTECH INC.  
 XX  
 Wu TD, Zhang Z, Zhou Y;  
 PI  
 WPI; 2004-347921/32.  
 DR  
 New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 XX  
 Claim 1; SEQ ID NO 4236; 7273pp; English.  
 XX  
 The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be

CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT nucleic acid of the invention

XX SQ Sequence 2177 BP; 703 A; 363 C; 442 G; 669 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 13; Length 2177;

Best Local Similarity 83.3%; Pred. No. 6e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24

|||||  
Db 750 AAAAACTGGAATGTAAGGCTGTG 773

RESULT 20

ADRO7065/c

ID ADRO7065 standard; cDNA; 3081 BP.

XX AC ADRO7065;

XX DT 04-NOV-2004 (first entry)

XX DE Full length human cDNA useful for treating neurological disease Seq 571.

XX KW Gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;

XX KW osteoporosis; neurological disease; Alzheimer's disease;

XX KW Parkinson's disease; dementia; short memory; cancer;

XX KW sense or motor function; emotional reaction; fear response; panic;

XX KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytotstatic;

XX KW tranquiliser.

XX OS Homo sapiens.

XX EP1447413-A2.

XX PD 18-AUG-2004.

XX PF 12-FEB-2004; 2004EP-000031145.

XX PR 14-FEB-2003; 2003JP-00102207.

XX PR 09-MAY-2003; 2003JP-00131452.

XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;

XX PI Wakamatsu A, Ishii S, Nagai K, Irie R;

XX DR WPI: 2004-583265/57.

XX DR P-PSDB; ADRO9021.

XX PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX PS Claim 1; SEQ ID NO 571; 2686pp; English.

XX CC This invention relates to novel, isolated full length human cDNA  
CC molecules and the encoded proteins thereof. Specifically, it refers to  
CC cDNA clones obtained by an oligo-capping method, where none of these  
CC clones are identical to any known human mRNAs. The present invention  
CC describes an immunoassay to identify agonists and antagonists, as well as  
CC antibodies, antisense molecules and siRNAs that can all be used to bind  
CC to and modulate expression of the cDNA molecules. As such, these  
CC molecules are useful for diagnostic markers or therapeutic targets for  
CC the various diseases or morbid states. In particular, they are useful in  
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
CC disease, Parkinson's disease, dementia, short memory and various cancers,  
CC as well as for maintaining equilibrium of sense or motor function, and  
CC for treating emotional reaction, fear response and panic. Accordingly,  
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
CC cytotstatic and tranquiliser activities. This polynucleotide is a full  
CC length human cDNA sequence of the invention. NOTE: This sequence is not  
CC given in the sequence listing of the specification but can be obtained on

CC CD-ROM from the European Patent Office, Vienna Sub-office.

XX SQ Sequence 3081 BP; 861 A; 613 C; 682 G; 925 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 13; Length 3081;

Best Local Similarity 83.3%; Pred. No. 6.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25

|||||  
Db 401 AAAAACTGGAATCAGGCTGACA 378

RESULT 21

ABX63354/c

ID ABX63354 standard; cDNA; 5271 BP.

XX AC ABX63354;

XX DT 25-FEB-2003 (first entry)

XX DE Human cDNA #354 differentially expressed in activated vascular tissue.

XX KW Human; gene; ss; vascular tissue; cytotstatic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.

XX OS Homo sapiens.

XX US2002137081-A1.

XX PD 26-SEP-2002.

XX PF 08-JAN-2002; 2002US-00044090.

XX PR 28-JUL-2000; 2000US-0222469P.

XX PR 08-JAN-2001; 2001US-0260483P.

XX PA (BAND/) BANDMAN O.

XX PI Bandman O;

XX DR WPI; 2003-110597/10.

XX PT Combination for diagnosing, staging, treating, or monitoring the  
PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
PT comprises several cDNAs that are differentially expressed in activated  
PT vascular tissue.

XX PS Claim 1; Page; 18pp; English.

XX CC This invention relates to a combination comprising several cDNAs that are  
CC differentially expressed in activated vascular tissue. The invention also  
CC discloses a high throughput method for detecting differentially expressed  
CC cDNAs in a sample. The cDNAs of the invention may have  
CC antiatherosclerotic; cytotstatic; cardiant; hypotensive; antidiabetic;  
CC gynaecological; vasotropic and cerebroprotective activities and may be  
CC used in gene therapy. The cDNAs of the invention may be used in a high-  
CC throughput methods for detecting differential expression of one or more  
CC cDNAs in a sample, or screening several molecules or compounds to  
CC identify a molecule or compound that specifically binds a cDNA of the  
CC invention. A protein encoded by the cDNA may be used to screen several  
CC molecules or compounds to identify a ligand that specifically binds to  
CC the protein, or to produce or purify an antibody to the protein that can  
CC be used to detect a protein in a sample or purify a natural or  
CC recombinant protein from a sample. The nucleotides may be useful for  
CC diagnosing, staging, treating, or monitoring the progression of treatment  
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion  
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
CC genetic or gene expression analysis of several new nucleic acid

CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for  
CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
CC associated with abnormalities in the expression, amount or distribution  
CC of the protein. The present sequence represents a cDNA of the invention  
CC that is differentially expressed in activated vascular tissue. Note: The  
CC sequence data for this patent did not form part of the specification, but  
CC was obtained in electronic format directly from USPTO at  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20020137081>  
XX  
SQ Sequence 5271 BP; 1374 A; 1122 C; 1141 G; 1294 T; 0 U; 340 Other;

Query Match 70.4%; Score 17.6; DB 8; Length 5271;  
Best Local Similarity 83.3%; Pred. No. 6.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
|||||  
Db 3345 AAAAATCTGGGATTACAGGCTGAG 3322

RESULT 22  
ADJ56534/c  
ID ADJ56534 standard; cDNA; 5271 BP.

AC ADJ56534;

XX 06-MAY-2004 (first entry)

DT Rat cDNA differentially expressed in MYCN activated cells SeqID 340.

DE rat; differential expression; transactivator; proto-oncogene;  
KW neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;  
KW MYCN activated cell.

XX Rattus norvegicus.

OS US2003119009-A1.

PN 26-JUN-2003.

XX 25-FEB-2002; 2002US-00084817.

XX 23-FEB-2001; 2001US-0270784P.

XX (STUA/) STUART S G.

PA (NUCH/) NUCHTERN J G.

PA (PLON/) PLON S E.

PA (SHOH/) SHOHET J M.

XX Stuart SG, Nuchtern JG, Plon SE, Shohet JM;

XX WPI; 2003-635698/60.

XX New genes regulated by MYCN activation, useful in gene therapy.  
PT Particularly for treating a subject with e.g. neuroblastoma or other  
PT cancers, or for diagnosing, staging or monitoring the treatment of the  
PT cancer.

PS Claim 1; SEQ ID NO 340; 27pp; English.

XX This invention relates to novel isolated cDNAs that are differentially  
CC expressed in MYCN activated cells. Specifically, it refers to  
CC polynucleotide sequences that exhibit differential expression patterns in  
CC cells activated by the transactivator MYCN, where MYCN is a proto-  
CC oncogene that is amplified in neuroblastoma cells and is common in small  
CC cell lung cancers. The present invention describes these cDNA molecules  
CC as useful for in hybridisation assays to detect expression of nucleic  
CC acids (or complementary nucleic acids) in a present in a given sample, as  
CC well as for screening assays by identifying molecules or compounds that  
CC specifically bind the cDNA as a ligand and modulate function or activity.  
CC Accordingly, these compositions exhibit cytostatic activity and can also  
CC be used for gene therapy purposes. This polynucleotide sequence is a cDNA  
CC that is differentially expressed in MYCN activated cells, given in an

CC exemplification of the invention. NOTE: This sequence does not appear in  
CC the printed specification but has been obtained in electronic format from  
CC the US Patent Office at  
CC <ftp://seqdata.uspto.gov/sequence.html?DocID=20030119009>.

XX Sequence 5271 BP; 1374 A; 1122 C; 1141 G; 1294 T; 0 U; 340 Other;

Query Match 70.4%; Score 17.6; DB 10; Length 5271;  
Best Local Similarity 83.3%; Pred. No. 6.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
|||||  
Db 3345 AAAAATCTGGGATTACAGGCTGAG 3322

RESULT 23  
ABS76688  
ID ABS76688 standard; DNA; 14815 BP.

XX ABS76688;

XX 12-DEC-2002 (first entry)

XX PRRS virus DNA sequence.

XX Porcine reproductive and respiratory syndrome; PRRS; virus;

XX open reading frame; ORF; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7;

XX PRRS infection; virucide; vaccine; pig; ds.

XX Porcine reproductive and respiratory syndrome virus.

XX WO200272802-A2.

XX 19-SEP-2002.

XX 07-MAR-2002; 2002WO-EP002486.

XX 09-MAR-2001; 2001US-0274603P.

XX (BOEH ) BOEHRINGER INGELHEIM VETMEDICA GMBH.

XX Elbers K, Peschke S, Schuetz B;

XX WPI; 2002-723343/78.

XX New attenuated European porcine reproductive and respiratory syndrome  
PT (PRRS) virus, useful for manufacturing a vaccine against PRRS, i.e. for  
PT the prophylaxis and treatment of PRRS infection.

XX Claim 6; Page 45-50; 75pp; English.

XX The present invention relates to a new attenuated European porcine  
CC reproductive and respiratory syndrome (PRRS) virus encoded by a nucleic  
CC acid comprising open reading frame (ORF)1, ORF2, ORF3, ORF4, ORF5, ORF6  
CC and ORF7. The attenuated European PRRS virus is useful in the manufacture  
CC of a vaccine against PRRS, i.e. for the prophylaxis and treatment of PRRS  
CC infection. The present nucleic acid sequence represents the PRRS virus  
CC sequence of the invention

XX Sequence 14815 BP; 3184 A; 4082 C; 3717 G; 3830 T; 0 U; 2 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 14815;  
Best Local Similarity 83.3%; Pred. No. 8.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
Db 9965 AAAAATCTGAATCTAAGGCTAGGA 9988

RESULT 24  
AD213631\_4/c

```
Continuation (5 of 5) of ADZ13631 from base 40001 (Human cancer-associated genomic DNA
WP Sequence split into 5 fragments LOCUS ADZ13631 Accession Adz13631
WP Fragment Name Begin End
WP ADZ13631_0 1 110000
WP ADZ13631_1 100001 210000
WP ADZ13631_2 200001 310000
WP ADZ13631_3 300001 410000
WP ADZ13631_4 400001 420555

Query Match 70.4%; Score 17.6; DB 14; Length 20555;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||||| |||
Db 18583 AAAAACTGAAAAATCAGGCTTAGA 18560

RESULT 25
ADZ13620_4/c
Continuation (5 of 5) of ADZ13620 from base 40001 (Human cancer-associated genomic DNA
WP Sequence split into 5 fragments LOCUS ADZ13620 Accession Adz13620
WP Fragment Name Begin End
WP ADZ13620_0 1 110000
WP ADZ13620_1 100001 210000
WP ADZ13620_2 200001 310000
WP ADZ13620_3 300001 410000
WP ADZ13620_4 400001 420555

Query Match 70.4%; Score 17.6; DB 14; Length 20555;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||||| |||
Db 18583 AAAAACTGAAAAATCAGGCTTAGA 18560

RESULT 26
AAD43979/c
ID AAD43979 standard; DNA; 24526 BP.
XX
AC AAD43979;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human transporter gene.
XX
KW Human; transporter protein; transgenic; pharmacogenomic analysis;
KW Gene therapy; chromosome 1; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(48,G)
FT /tag= a
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(132,A)
FT /tag= b
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(724,C)
FT /tag= c
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(1558,G)
FT /tag= d
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(1577,G)
FT /tag= e
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT CDS 1997..22526
FT /tag= f
FT /product= "Human transporter protein"
FT exon 1997..2121
```

```
FT intron /tag= g
FT 2122..4732
FT /tag= h
FT variation replace(2487,A)
FT /tag= i
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(2634,C)
FT /tag= j
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(4352,G)
FT /tag= k
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT exon 4733..4872
FT /tag= l
FT intron 4873..5004
FT /tag= m
FT exon 5005..5115
FT /tag= n
FT intron 5116..5781
FT /tag= o
FT variation replace(5157,C)
FT /tag= p
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(5658,T)
FT /tag= q
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT exon 5782..5957
FT /tag= r
FT variation replace(5945,C)
FT /tag= s
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT intron 5958..7770
FT /tag= t
FT variation replace(6281,C)
FT /tag= u
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(6452,C)
FT /tag= v
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(6610,T)
FT /tag= w
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(7247,C)
FT /tag= x
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(7360,G)
FT /tag= y
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(7644,T)
FT /tag= z
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT exon 7771..7935
FT /tag= aa
FT intron 7936..8470
FT /tag= ab
FT variation replace(8127,C)
FT /tag= ac
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(8317,A)
FT /tag= ad
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT exon 8471..8623
FT /tag= ae
FT intron 8624..8917
FT /tag= af
FT exon 8918..9000
FT /tag= ag
FT intron 9001..9777
FT /tag= ah
FT variation replace(9079,A)
FT /tag= ai
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT variation replace(9537,G)
```





```
FT      /*tag= j  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(4352,G)  
FT      /*tag= k  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      4733. .4872  
FT      /*tag= l  
FT      4873. .5004  
FT      /*tag= m  
FT      5005. .5115  
FT      /*tag= n  
FT      5116. .5781  
FT      /*tag= o  
FT      replace(5157,C)  
FT      /*tag= p  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(5658,T)  
FT      /*tag= q  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      5782. .5957  
FT      /*tag= r  
FT      replace(5945,C)  
FT      /*tag= s  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      5958. .7770  
FT      /*tag= t  
FT      replace(6281,C)  
FT      /*tag= u  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(6452,C)  
FT      /*tag= v  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(6610,T)  
FT      /*tag= w  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(7247,C)  
FT      /*tag= x  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(7360,G)  
FT      /*tag= y  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(7644,T)  
FT      /*tag= z  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      7771. .7935  
FT      /*tag= aa  
FT      7936. .8470  
FT      /*tag= ab  
FT      replace(8127,C)  
FT      /*tag= ac  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(8317,A)  
FT      /*tag= ad  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      8471. .8623  
FT      /*tag= ae  
FT      8624. .8917  
FT      /*tag= af  
FT      8918. .9000  
FT      /*tag= ag  
FT      9001. .9777  
FT      /*tag= ah  
FT      replace(9079,A)  
FT      /*tag= ai  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(9537,G)  
FT      /*tag= aj  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      9778. .9925  
FT      /*tag= ak  
FT      9926. .10221  
FT      /*tag= al  
FT      10222. .10335
```

```
FT      intron  
FT      /*tag= am  
FT      10336. .10539  
FT      /*tag= an  
FT      10540. .10617  
FT      /*tag= ao  
FT      10618. .11197  
FT      /*tag= ap  
FT      11198. .11293  
FT      /*tag= aq  
FT      11294. .13338  
FT      /*tag= ar  
FT      replace(12302,G)  
FT      /*tag= as  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(12354,C)  
FT      /*tag= at  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(12487,C)  
FT      /*tag= au  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(13197. .13199,CA)  
FT      /*tag= av  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(13257,G)  
FT      /*tag= aw  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      13339. .13445  
FT      /*tag= ax  
FT      13446. .14214  
FT      /*tag= ay  
FT      14215. .14284  
FT      /*tag= az  
FT      14285. .14400  
FT      /*tag= ba  
FT      14401. .14493  
FT      /*tag= bb  
FT      14494. .15980  
FT      /*tag= bc  
FT      replace(14541,A)  
FT      /*tag= bd  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(14545,G)  
FT      /*tag= be  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(15041,A)  
FT      /*tag= bf  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(15053,C)  
FT      /*tag= bg  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(15065,G)  
FT      /*tag= bh  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(15108,C)  
FT      /*tag= bi  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      15981. .16262  
FT      /*tag= bj  
FT      16263. .17597  
FT      /*tag= bk  
FT      replace(16273. .16275,CG)  
FT      /*tag= bl  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      replace(17424,T)  
FT      /*tag= bm  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      17598. .17652  
FT      /*tag= bn  
FT      replace(17627,A)  
FT      /*tag= bo  
FT      /standard_name= "Single nucleotide polymorphism (SNP) "  
FT      17653. .18842  
FT      /*tag= bp
```

```
FT variation replace(18427,C)
FT /tag= bq
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18813,G)
FT /tag= br
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 18843..18988
FT /tag= bs
FT 18989..20477
FT /tag= bt
FT replace(19035,C)
FT /tag= bu
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(19182,C)
FT
FT
FT Query Match 70.4%; Score 17.6; DB 13; Length 24526;
FT Best Local Similarity 83.3%; Pred. No. 8.9e+02;
FT Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
FT
FT QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
FT |||||
FT 3296 AAGAACTGGAAACAGAGGCTGAGA 3273
FT
FT
FT RESULT 28
FT AB224431/C
FT ID AB224431 standard; cDNA; 26320 BP.
FT XX AC
FT XX AB224431;
FT
FT DT 18-MAR-2003 (first entry)
FT
FT DE Human constitutively active receptor (hCAR) gene.
FT
FT XX
FT KW Constitutively active reactor; human; hCAR; receptor;
FT G-protein coupled receptor; cytostatic; neuroleptic; antiparkinsonian;
FT KW single nucleotide polymorphism; SNP; chromosome 4; gene therapy; gene;
FT KW db.
FT
FT XX Homo sapiens.
FT
FT FH Location/Qualifiers
FT FH variation replace(469,C)
FT /tag= h
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221921, position 8317095 on chromosome 4"
FT variation replace(469,C)
FT /tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221921, position 8317095 on chromosome 4"
FT variation replace(568,A)
FT /tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221920, position 8317194 on chromosome 4"
FT variation replace(865,C)
FT /tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221919, position 8317491 on chromosome 4"
FT variation replace(1046,T)
FT /tag= k
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221918, position 8317672 on chromosome 4"
FT variation replace(1637,T)
FT /tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221917, position 8318263 on chromosome 4"
FT variation replace(2349,G)
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221916, position 8319575 on chromosome 4"
FT variation replace(3335,C)
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT
FT
FT /note= "SNP CV1221915, position 8319961 on chromosome 4"
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV7662683, position 8321056 on chromosome 4"
FT replace(6658,T)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221914, position 8323284 on chromosome 4"
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221913, position 8324706 on chromosome 4"
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221912, position 8324712 on chromosome 4"
FT replace(8090,C)
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221911, position 8324716 on chromosome 4"
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221910, position 8325253 on chromosome 4"
FT replace(9804,G)
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221909, position 8326430 on chromosome 4"
FT 10001..16380
FT /tag= a
FT /product= "Human constitutively active receptor"
FT /note= "contains introns"
FT 10001..10668
FT /tag= b
FT 10669..11548
FT /tag= c
FT 11549..11662
FT /tag= d
FT 11663..16073
FT /tag= e
FT replace(13860,T)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221908, position 8330486 on chromosome 4"
FT replace(13889,G)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221907, position 8330515 on chromosome 4"
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221906, position 8331272 on chromosome 4"
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221905, position 8331384 on chromosome 4"
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221904, position 8331760 on chromosome 4"
FT replace(15208,-)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV7664553, position 8331834 on chromosome 4"
FT replace(15253,G)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP CV1221902, position 8331879 on chromosome 4"
FT 16074..16383
FT /tag= f
FT replace(16279,A)
FT /tag= ac
FT variation
```

FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221901, position 8332905 on chromosome 4"  
FT replace(16448,T)  
FT /\*tag= ad  
FT  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221900, position 8333074 on chromosome 4"  
FT replace(16499,G)  
FT /\*tag= ae  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221899, position 8333125 on chromosome 4"  
FT replace(16534,T)  
FT /\*tag= af  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221898, position 8333160 on chromosome 4"  
FT replace(19520,G)  
FT /\*tag= ag  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221897, position 8336146 on chromosome 4"  
FT replace(19647,C)  
FT /\*tag= ah  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221896, position 8336273 on chromosome 4"  
FT replace(19999,T)  
FT /\*tag= ai  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221895, position 8336625 on chromosome 4"  
FT replace(20807,T)  
FT /\*tag= ak  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221893, position 8337433 on chromosome 4"  
FT replace(23179,A)  
FT /\*tag= al  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221892, position 8339805 on chromosome 4"  
FT replace(23224,T)  
FT /\*tag= am  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221891, position 8339850 on chromosome 4"  
FT replace(23251,A)  
FT /\*tag= an  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221890, position 8339877 on chromosome 4"  
FT replace(23666,-)  
FT /\*tag= ao  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV7664594, position 8340292 on chromosome 4"  
FT replace(23671,G)  
FT /\*tag= ap  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV7664595, position 8340297 on chromosome 4"  
FT replace(23770,A)  
FT /\*tag= aq  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221887, position 8340396 on chromosome 4"  
FT replace(23795,-)  
FT /\*tag= ar  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV7664734, position 8340421 on chromosome 4"  
FT replace(24037,A)  
FT /\*tag= as  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221885, position 8340663 on chromosome 4"  
FT replace(24431,C)  
FT /\*tag= at  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221884, position 8341057 on chromosome 4"  
FT replace(24556,C)  
FT /\*tag= au  
FT /standard\_name= "Single nucleotide polymorphism"  
FT /note= "SNP CV1221883, position 8341182 on chromosome 4"  
FT replace(29155,A)  
FT /\*tag= aj  
FT /standard\_name= "Single nucleotide polymorphism"

FT /note= "SNP CV1221894, position 8336781 on chromosome 4"  
XX WO2002101005-A2.  
PN  
XX  
PD 19-DEC-2002.  
XX  
XX 30-MAY-2002; 2002WO-US016834.  
PF  
  
Query Match 70.4%; Score 17.6; DB 10; Length 26320;  
Best Local Similarity 83.3%; Pred. No. 9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 AAAAACTGGAAATCTCAGGCTGAG 24  
||||| ||||| ||||| ||||| ||  
Db 4864 AAAAACCTGGAAATGCCAGGCTAAG 4841  
  
RESULT 29  
AD259528/c  
ID AD259528 standard; DNA; 37442 BP.  
XX  
AC AD259528;  
XX  
DT 30-JUN-2005 (first entry)  
XX  
DE Secondary hyperparathyroidism detection polymorphic gene IL12RB1, SEQ 22.  
XX  
KW secondary hyperparathyroidism; endocrine-gen.; antithyroid;  
KW renal failure; nephrotropic; SNP detection;  
KW single nucleotide polymorphism; SNP; gene; ds; IL12RB1.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH variation replace(27442,A)  
FT /\*tag= a  
FT /standard\_name= "Single nucleotide polymorphism"  
XX  
XX JP2005102601-A.  
PN  
XX 21-APR-2005.  
XX  
XX 30-SEP-2003; 2003JP-00341015.  
PF  
XX 30-SEP-2003; 2003JP-00341015.  
PR  
XX (HYUB-) HYUBITTO GENOMICS KK.  
XX (JIKE-) UNIV JIKEI.  
XX  
XX WPI; 2005-358641/37.  
DR  
XX  
XX Testing secondary hyperparathyroidism in chronic renal failure patient,  
PT involves detecting variation in gene chosen from CACNA1C, CALCL, CH13L1,  
PT EGF, FGF1, GFR1, GPR56 and GPRK6.  
XX  
XX Claim 4; SEQ ID NO 22; 138pp; Japanese.  
PS  
XX  
XX The invention relates to a novel method for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. The method  
CC involves detecting a variation in a gene chosen from CACNA1C, CALCL,  
CC CH13L1, EGF, FGF1, GFR1, GPR56, GPRK6, IL10RA, IL12RB1, KCNJ14,  
CC KCNQ1, ORCT14, PDGFRA, SCYB14, SLC12A1, SLC2A3, TGFBR3, TMEM1, CALCR,  
CC IL17R, OSTF1, FGF6, HGF, MET, TGFBI and VEGF, or detecting the base in a  
CC polymorphism region existing in the vicinity of any one of the genes. The  
CC invention further comprises a reagent or kit for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. This  
CC polynucleotide sequence represents the polymorphism containing human  
CC IL12RB1 gene of the invention.  
XX  
XX Sequence 37442 BP; 9374 A; 9810 C; 9241 G; 9017 T; 0 U; 0 Other;  
SQ  
  
Query Match 70.4%; Score 17.6; DB 14; Length 37442;  
Best Local Similarity 83.3%; Pred. No. 9.5e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAG 24  
|||||  
Db 18487 AAAATACTGGAATTACAGGCTGAG 18464

## RESULT 30

ID ADZ59517/C  
ADZ59517 standard; DNA; 37442 BP.

XX AC ADZ59517;

XX 30-JUN-2005 (first entry)

XX Secondary hyperparathyroidism detection polymorphic gene IL12RB1, SEQ 11.

XX secondary hyperparathyroidism; endocrine-gen.; antithyroid;

KW renal failure; nephrotropic; SNP detection;

KW single nucleotide polymorphism; SNP; gene; ds; IL12RB1.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT variation replace(10001,G)

FT /\*tag= a

FT /standard\_name= "Single nucleotide polymorphism"

XX JP2005102601-A.

XX 21-APR-2005.

XX 30-SEP-2003; 2003JP-00341015.

XX 30-SEP-2003; 2003JP-00341015.

XX (HYUB-) HYUBITTO GENOMICS KK.

XX (JIKI-) UNIV JIKEI.

XX WPI; 2005-358641/37.

XX Testing secondary hyperparathyroidism in chronic renal failure patient,  
PT involves detecting variation in gene chosen from CACNA1C, CALCL, CHI3L1,  
PT EGF, FGF1, GFRA1, GPR56 and GPRK6.

XX Claim 4; SEQ ID NO 11; 138pp; Japanese.

XX The invention relates to a novel method for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. The method  
CC involves detecting a variation in a gene chosen from CACNA1C, CALCL,  
CC CHI3L1, EGF, FGF1, GFRA1, GPR56, GPRK6, IL10RA, IL10RB, IL12RB1, KCNJ14,  
CC KCNQ1, ORCTL4, PDGFRA, SCYB14, SLC12A1, SLC2A3, TGFBR3, TMEM1, CALCR,  
CC IL17R, OSTF1, FGF6, HGF, MET, TGFBI and VEGF, or detecting the base in a  
CC polymorphism region existing in the vicinity of any one of the genes. The  
CC invention further comprises a reagent or kit for testing secondary  
CC hyperparathyroidism in a chronic renal failure patient. This  
CC polymucleotide sequence represents the polymorphism containing human  
CC IL12RB1 gene of the invention.

XX SQ Sequence 37442 BP; 9374 A; 9810 C; 9241 G; 9017 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 14; Length 37442;

Best Local Similarity 83.3%; Pred. No. 9.5e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAG 24  
|||||

Db 18487 AAAATACTGGAATTACAGGCTGAG 18464

## RESULT 31

AEB96540/C

ID AEB96540 standard; DNA; 37442 BP.

XX AEB96540;

XX 06-OCT-2005 (first entry)

XX Human IL12RB1 gene, SEQ ID 24.

XX hepatitis C virus infection; antiinflammatory; hepatotropic; virucide;

KW liver cirrhosis; fibrosis; hepatoma; SNP detection; IL12RB1; ds.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT variation 107

FT /\*tag= a

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 109

FT /\*tag= b

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 191

FT /\*tag= c

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 769

FT /\*tag= d

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 1160

FT /\*tag= e

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 1693

FT /\*tag= f

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 3410

FT /\*tag= g

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 5893

FT /\*tag= h

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 6662

FT /\*tag= i

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 6929

FT /\*tag= j

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 6989

FT /\*tag= k

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 7715

FT /\*tag= l

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 8922

FT /\*tag= m

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 9091

FT /\*tag= n

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 9879

FT /\*tag= o

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 10001

FT /\*tag= p

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 10043

FT /\*tag= q

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 10072

FT /\*tag= r

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 10220

FT /\*tag= s

FT /standard\_name= "Single nucleotide polymorphism"

FT variation 10258

FT /\*tag= t

FT /standard\_name= "Single nucleotide polymorphism"

FT	variation	10453	FT	single nucleotide polymorphism
FT	variation	10671	FT	single nucleotide polymorphism
FT	variation	10817	FT	single nucleotide polymorphism
FT	variation	11085	FT	single nucleotide polymorphism
FT	variation	11376	FT	single nucleotide polymorphism
FT	variation	11586	FT	single nucleotide polymorphism
FT	variation	11757	FT	single nucleotide polymorphism
FT	variation	12073	FT	single nucleotide polymorphism
FT	variation	12220	FT	single nucleotide polymorphism
FT	variation	12317	FT	single nucleotide polymorphism
FT	variation	12578	FT	single nucleotide polymorphism
FT	variation	13297	FT	single nucleotide polymorphism
FT	variation	13336	FT	single nucleotide polymorphism
FT	variation	13724	FT	single nucleotide polymorphism
FT	variation	14454	FT	single nucleotide polymorphism
FT	variation	14999	FT	single nucleotide polymorphism
FT	variation	15258	FT	single nucleotide polymorphism
FT	variation	15460	FT	single nucleotide polymorphism
FT	variation	16176	FT	single nucleotide polymorphism
FT	variation	16559	FT	single nucleotide polymorphism
FT	variation	17369	FT	single nucleotide polymorphism
FT	variation	17425	FT	single nucleotide polymorphism
FT	variation	17706	FT	single nucleotide polymorphism
FT	variation	18110	FT	single nucleotide polymorphism
FT	variation	18488	FT	single nucleotide polymorphism

```

Query Match          70.4%; Score 17.6; DB 14; Length 37442;
Best Local Similarity 83.3%; Pred. No. 9.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAACTGGAATCTCAGGCTGAG 24
          ||||| ||||| ||||| |||||
Db      18487 AAAATACTGGAATTACAGGCTGAG 18464

RESULT 32
ABL07876
ID ID ABL07876 standard; cDNA; 37996 BP.
XX AC
XX ABL07876;
XX AC
XX XX
DT 26-MAR-2002 (first entry)
XX XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 18110.
XX XX

```





CC The nucleic acid molecules, proteins and antibodies can be used in  
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
CC and nucleic acid molecules, etc. are useful for production of transgenic  
CC animals, especially a transgenic mouse for the study of diseases  
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and  
CC achondroplasia. This sequence encodes the murine LOBO protein described  
CC in the method of the invention

SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 2; Length 49999;

Best Local Similarity 83.3%; Pred. No. 9.9e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25

DB 45957 AAAAACTGGAATCTCAGGCTGAGA 45934

RESULT 35

ADQ97081/C

ID ADQ97081 standard; DNA; 63411 BP.

XX

AC ADQ97081;

XX 07-OCT-2004 (first entry)

XX Mouse cancer associated sequence MD10-004, SEQ ID 57.

DE Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.

XX Mus musculus.

XX WO2004060304-A2.

XX 22-JUL-2004.

XX 22-DEC-2003; 2003WO-US041389.

XX 27-DEC-2002; 2002US-00330773.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-543781/52.

XX New isolated cancer associated nucleic acids comprising at least 10  
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
PT cancers such as leukemia and lymphoma.

XX Claim 1; SEQ ID NO 57; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-  
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 63411 BP; 17408 A; 15472 C; 14344 G; 15847 T; 0 U; 340 Other;

Query Match 70.4%; Score 17.6; DB 12; Length 63411;

Best Local Similarity 83.3%; Pred. No. 1e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24

DB 41743 AAAAACTGGAATCTCAGGCTGAG 41720

RESULT 36

ADC85257

ID ADC85257 standard; DNA; 96594 BP.

XX

AC ADC85257;

XX 01-JAN-2004 (first entry)

XX Mouse Ptpkr genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;

XX secreted; transmembrane; intracellular; ds.

XX Mus sp.

XX WO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of  
PT the carcinoma-associated (CA) genes, useful for screening for drug  
PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 43; 983pp; English.

XX The invention relates to a recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the fully defined carcinoma-  
CC associated (CA) genes from the 50 tables given in the specification. The  
CC CA proteins are secreted, transmembrane or intracellular proteins. The  
CC recombinant nucleic acids are useful for screening for drug candidates  
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-  
CC ADC85514 represent CA genes of the invention.

XX Sequence 96594 BP; 27993 A; 18068 C; 18732 G; 30838 T; 0 U; 963 Other;

Query Match 70.4%; Score 17.6; DB 10; Length 96594;

Best Local Similarity 83.3%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24

DB 27319 AAAAACTGGAATCTCAGGCTGAG 27342

RESULT 37

ADA02777

ID ADA02777 standard; DNA; 96595 BP.

XX

AC ADA02777;

XX 06-NOV-2003 (first entry)

XX Mouse Ptpkr carcinoma associated gene, SEQ ID NO:1295.

XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.

XX Mus sp.

XX WO2003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

```

XX 26-DEC-2001; 2001US-00035832.
PR (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX PI
XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1295; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed murine CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 96595 BP; 27994 A; 18068 C; 18732 G; 30838 T; 0 U; 963 Other;
SQ
Query Match 70.4%; Score 17.6; DB 9; Length 96595;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
DB 27319 ATAAATCTGTAATCTCAGGCTGAG 27342

RESULT 38
ADB72515
ID ADB72515 standard; DNA; 96595 BP.
XX
XX ADB72515;
AC
XX 04-DEC-2003 (first entry)
DT
XX Mouse Ptprk gene.
DE
XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX Mus sp.
OS
XX WO2003008583-A2.
XX
XX 30-JAN-2003.
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
XX
XX 23-OCT-2001; 2001US-00004113.
XX
XX 08-NOV-2001; 2001US-00052482.
XX
XX 30-NOV-2001; 2001US-00997722.

```

```

PR 20-DEC-2001; 2001US-00034650.
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
XX PI
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 343; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
XX Sequence 96595 BP; 27994 A; 18068 C; 18732 G; 30838 T; 0 U; 963 Other;
SQ
Query Match 70.4%; Score 17.6; DB 10; Length 96595;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
DB 27319 ATAAATCTGTAATCTCAGGCTGAG 27342

RESULT 39
ADM74372
ID ADM74372 standard; DNA; 96595 BP.
XX
XX ADM74372;
AC
XX 01-JUL-2004 (first entry)
DT
XX Murine carcinoma associated (CA) nucleic acid #22.
XX
XX Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
XX carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
XX cytostatic.
XX
XX Mus musculus.
OS
XX US2004072154-A1.
XX
XX 15-APR-2004.
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX 22-DEC-2000; 2000US-00747377.
XX
XX 02-MAR-2001; 2001US-00798586.
XX
XX (MORR/) MORRIS D W.
XX
XX (ENGE/) ENGELHARD E K.
XX
XX Morris DW, Engelhard EK;
XX
XX WPI; 2004-328562/30.
XX
XX New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
XX lymphoma.
XX
XX Claim 1; SEQ ID NO 43; 29pp; English.
XX
XX The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic

```

CC acid, a recombinant protein, a method of screening for drug candidates, a  
CC method of screening for a bioactive agent capable of binding to a  
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
CC method of screening for a bioactive agent capable of modulating the  
CC activity of a CAP, a method of evaluating the effect of a candidate  
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting  
CC the activity of a CAP, a method of treating carcinomas, a method of  
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
CC propensity to carcinoma. A method of evaluating the effect of a candidate  
CC carcinoma drug comprises administering the drug to a patient, removing a  
CC cell sample from the patient and determining alterations in the  
CC expression or activation of a gene comprising the nucleotide sequence. A  
CC method of diagnosing carcinoma comprises determining the expression of  
CC one or more genes comprising the nucleic acid sequence in a first tissue  
CC type of a first individual and comparing the expression of the gene from  
CC a second normal tissue type from the first individual or a second  
CC unaffected individual, where a difference in the expression indicates  
CC that the first individual has carcinoma. A method of inhibiting the  
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
CC carcinomas comprises administering to a patient an inhibitor of CAP.  
CC Neutralising the effect of a CAP comprises contacting an agent specific  
CC for the CAP. The polypeptide specifically binds to the protein encoded by  
CC the nucleic acid. It comprises an antibody that specifically binds to the  
CC protein encoded by the nucleic acid. The nucleic acids are useful for  
CC preparing a composition for diagnosing or treating carcinoma e.g.,  
CC leukaemia or lymphoma. This sequence represents a murine carcinoma  
CC associated (CA) nucleic acid of the invention. Note: The sequence data  
CC for this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

XX SQ Sequence 96595 BP; 27994 A; 18068 C; 18732 G; 30838 T; 0 U; 963 Other;

Query Match 70.4%; Score 17.6; DB 12; Length 96595;

Best Local Similarity 83.3%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24

Db 27319 ATAATCTGTAATCTCAGGCTGAG 27342

RESULT 40

ACN43998\_2/c  
Continuation (3 of 7) of ACN43998 from base 200001 (Human genomic sequence HCG1778483.)

WP Sequence split into 7 fragments LOCUS ACN43998 Accession ACN43998

WP	Fragment Name	Begin	End
WP	ACN43998_0	1	110000
WP	ACN43998_1	100001	210000
WP	ACN43998_2	200001	310000
WP	ACN43998_3	300001	410000
WP	ACN43998_4	400001	510000
WP	ACN43998_5	500001	610000
WP	ACN43998_6	600001	653122

Query Match 70.4%; Score 17.6; DB 11; Length 110000;

Best Local Similarity 83.3%; Pred. No. 1.1e+03;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25

Db 51517 AAAAATAAATCTCAGACTGAGA 51494

RESULT 41

ACN45090\_0  
WP Sequence split into 4 fragments LOCUS ACN45090 Accession ACN45090

WP	Fragment Name	Begin	End
WP	ACN45090_0	1	110000
WP	ACN45090_1	100001	210000
WP	ACN45090_2	200001	310000
WP	ACN45090_3	300001	350764

ID ACN45090 standard; DNA; 350764 BP.

XX ACN45090;  
AC 18-NOV-2004 (first entry)  
DT Human genomic sequence HCG22125.  
DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
KW Homo sapiens.  
OS WO2003073826-A2.  
PN 12-SEP-2003.  
PD 28-FEB-2003; 2003WO-US006235.  
PF 01-MAR-2002; 2002US-00087192.  
PR (SAGR-) SAGRES DISCOVERY.  
PA Morris DW;  
PI WPI; 2003-328604/31.  
DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
PT comprises a nucleotide sequence.  
XX Claim 1; SEQ ID NO 1864; Opp; English.  
XX The present invention relates to novel DNA and protein sequences which  
CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published  
XX SQ Sequence 350764 BP; 99569 A; 63584 C; 68015 G; 119596 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 11; Length 110000;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
Db 60025 AAAATGCTGCATCTCTCAGGCTGAG 60048  
RESULT 42  
ABQ88198  
ID ABQ88198 standard; cDNA; 154902 BP.  
AC ABQ88198;  
XX 18-SEP-2002 (first entry)  
DE Human osteoblast differentiation related cDNA SEQ ID NO 105.  
KW Human; osteoblast; stem cell differentiation; bone tissue deposition;  
KW osteoporosis; osteopathic; ss.  
OS Homo sapiens.  
XX WO200250301-A2.  
XX

PD 27-JUN-2002.  
XX  
PF 18-DEC-2001; 2001WO-US048276.  
XX  
XX 18-DEC-2000; 2000US-0255882P.  
PR 24-APR-2001; 2001US-0285691P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX (PROC ) PROCTER & GAMBLE CO.  
XX  
PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
PI Mertz L;  
XX  
DR WPI; 2002-557663/59.  
XX  
XX Use of genes and their expression profiles associated with osteoblast  
PT differentiation for screening modulators bone formation, for diagnosing  
PT or treating e.g. osteoporosis, or as markers for the differentiation  
PT process.  
XX  
PS Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.  
XX  
XX The invention relates to genes and their expression profiles are used  
CC for: (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
CC deposition of bone tissue, abnormal rate of osteoblast formation or  
CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
CC cited in (b), or monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
CC induced abnormalities in bone formation or bone loss, conditions that  
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
CC or fibrous dysplasia. The present sequence is that of an osteoblast  
CC differentiation associated cDNA marker of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;  
XX  
Query Match 70.4%; Score 17.6; DB 6; Length 154902;  
Best Local Similarity 83.3%; Pred. No. 1.2e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
XX  
QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
DB 33872 ATAATCTGGATCTCAGGCTGAGA 33895  
XX  
RESULT 43  
ADZ12614  
ID ADZ12614 standard; DNA; 289106 BP.  
XX  
AC ADZ12614;  
XX  
XX 16-JUN-2005 (first entry)  
XX  
XX Murine cancer-associated genomic DNA #13.  
XX  
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;  
XX Cytostatic; gene; ds.  
XX  
XX Mus sp.  
XX  
XX WO2005031001-A2.  
XX  
XX 07-APR-2005.  
XX  
XX 23-SEP-2004; 2004WO-US031617.  
XX  
XX 23-SEP-2003; 2003US-00669920.  
XX

PA (CHIR ) CHIRON CORP.  
XX  
PI Morris DW, Malandro MS;  
XX  
DR WPI; 2005-273395/28.  
XX  
XX Nucleic acid array useful for detecting cancer associated nucleic acid,  
PT comprises two or more nucleic acid probes.  
XX  
XX Disclosure; SEQ ID NO 134; 198pp; English.  
XX  
XX The invention relates to a nucleic acid array for detecting a cancer  
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
CC The invention also relates to a peptide array comprising two or more  
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
CC that binds to a polypeptide, an isolated antibody or its fragment which  
CC binds to a polypeptide, which is prepared by immunizing a host animal  
CC with a composition comprising the polypeptide or its antigen binding  
CC fragment and collecting cells from the host expressing antibodies against  
CC the antigen or its antigen binding fragment, a composition comprising the  
CC antibody and a carrier, a method of screening for anticancer activity, a  
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
CC method of treating cancer and a method of inhibiting expression of a CA  
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
CC nucleic acids. The antibody is useful for detecting the presence or  
CC absence of cancer cells in an individual which involves contacting cells  
CC from the individual with the antibody and detecting a complex of a CA  
CC protein from the cancer cells and the antibody, where the detection of  
CC the complex correlates with the presence of cancer cells in the  
CC individual. The composition is useful for inhibiting growth of cancer  
CC cells in an individual or for delivering a therapeutic agent to cancer  
CC cells in an individual. The invention is also useful for diagnosing  
CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
CC a cell. This sequence represents murine cancer-associated genomic DNA of  
CC the invention.  
XX  
SQ Sequence 289106 BP; 84569 A; 54316 C; 55029 G; 92921 T; 0 U; 2271 Other;  
XX  
Query Match 70.4%; Score 17.6; DB 14; Length 289106;  
Best Local Similarity 83.3%; Pred. No. 1.3e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
XX  
QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
DB 84553 ATAATCTGTAATCTCAGGCTGAG 84576  
XX  
RESULT 44  
ADQ59380/C  
ID ADQ59380 standard; DNA; 299598 BP.  
XX  
AC ADQ59380;  
XX  
XX 07-OCT-2004 (first entry)  
XX  
XX Human cancer-associated (CA) gene sequence SEQ ID NO:16.  
XX  
XX human; cancer-associated gene; cancer-associated protein; cytostatic;  
XX gene therapy; vaccine; tyrosine kinase antagonist;  
XX G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2004058288-A1.  
XX  
XX 15-JUL-2004.  
XX  
XX 15-DEC-2003; 2003WO-US040082.  
XX  
XX 17-DEC-2002; 2002US-00322696.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX

PI Morris DW, Malandro MS;  
XX WPI: 2004-543349/52.  
DR P-PSDB; ADQ59382.  
XX  
PT New cancer-associated nucleic acid for diagnosing, preventing or treating  
PT cancer (e.g. lymphoma) or for screening agents that may be used for  
PT treating or preventing cancer.  
XX  
XX Claim 16; SEQ ID NO 16; 143pp; English.  
XX  
CC The present invention describes human cancer-associated (CA) nucleotide  
CC sequences (1). Also described: (1) an expression vector comprising (1);  
CC (2) a host cell comprising (1) or the expression vector; (3) a microarray  
CC for detecting a CA nucleic acid; (4) an isolated polypeptide encoded  
CC within an open reading frame of a CA sequence; (5) an isolated antibody,  
CC or its antigen binding fragment, that binds to the above polypeptide; (6)  
CC a hybridoma that produces the monoclonal antibody described above; (7) a  
CC pharmaceutical composition comprising the antibody and a pharmaceutical  
CC excipient; (8) a kit for detecting or diagnosing cancer cells, comprising  
CC the above (monoclonal) antibody or polynucleotide that selectively  
CC hybridizes to any of the polynucleotide sequences mentioned above; (9)  
CC methods for diagnosing cancer or for detecting the presence or absence of  
CC cancer cells in an individual; (10) a method for inhibiting growth of  
CC cancer cells in an individual; (11) a method for delivering a therapeutic  
CC agent to cancer cells in an individual; (12) an electronic library  
CC comprising the polynucleotide or polypeptide, or their fragments;  
CC mentioned above; (13) a method of screening for anticancer activity; (14)  
CC methods for detecting cancer associated with expression of a polypeptide  
CC or the presence of the antibody in a test cell or serum sample; (15) a  
CC method for screening for a bioactive agent capable of modulating the  
CC activity of a CA protein encoded by the above nucleic acid molecule; and  
CC (16) a method for treating cancers. (1) has cytostatic activity, and can  
CC be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,  
CC and as a G-protein coupled receptor antagonist. The compositions and  
CC methods of the present invention can be used for diagnosing, preventing  
CC and treating cancer, especially lymphomas. They may also be used in  
CC screening for agents that may be used for treating or preventing cancer.  
CC The present sequence represents a human CA gene sequence, which is given  
CC in the exemplification of the present invention. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 299598 BP; 92218 A; 54563 C; 55801 G; 95585 T; 0 U; 1431 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 12; Length 299598;  
Best Local Similarity 83.3%; Pred. No. 1.3e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAAAACTGGAAATCTCAGGCTGAGA 25  
||||||| || ||||| |||  
Db 297755 AAAAACTGAAAAATCAGGCTTAGA 297732  
RESULT 45  
AD081499  
ID AD081499 standard; cDNA; 1059 BP.  
XX  
XX AD081499;  
XX  
XX 21-APR-2005 (first entry)  
XX  
XX Plant full length insert polynucleotide seqid 219.  
DE  
DE plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.

XX Unidentified.  
XX US2004034888-A1.  
XX  
XX 19-FEB-2004.  
XX  
XX 28-APR-2003; 2003US-00425114.  
XX  
XX 06-MAY-1999; 99US-00304517.  
XX 05-NOV-2001; 2001US-00985678.  
XX  
XX (LIUJ/) LIU J.  
XX (ZHOU/) ZHOU Y.  
XX (KOVA/) KOVALIC D K.  
XX (SCRE/) SCREEN S E.  
XX (TABAR/) TABASKA J E.  
XX (CAOY/) CAO Y.  
XX  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI: 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
XX pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
XX Claim 1; SEQ ID NO 219; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
XX polynucleotide consisting of a sequence encoding an amino acid sequence  
XX available in electronic form from the US patent office at  
XX ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
XX of the invention are also useful in physical arrays of molecules and as  
XX plant breeding markers. The recombinant DNA construct is useful for  
XX improving plant tolerance to cold, heat, drought, herbicides, extreme  
XX osmotic conditions, pathogens or pests, for manipulating growth rate in  
XX plant cells by modification of the cell cycle pathway, for conferring  
XX increased resistance to plant disease, for producing galactomannan,  
XX lignin or plant growth regulators, for increasing the rate of homologous  
XX recombination in plants, for improving yield by modification of  
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
XX or by providing improved plant growth and development under at least one  
XX stress condition or for modifying seed oil or protein yield and/or  
XX content. This sequence represents a plant full length insert  
XX polynucleotide that can be used in the recombinant DNA construct of the  
XX invention.  
XX  
XX Sequence 1059 BP; 323 A; 235 C; 266 G; 235 T; 0 U; 0 Other;  
SQ  
Query Match 69.6%; Score 17.4; DB 13; Length 1059;  
Best Local Similarity 94.7%; Pred. No. 6.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAAAACTGGAAATCTCAGGC 20  
||||||| ||||| ||||| |||||  
Db 49 AAAAGCTGGAATCTCAGGC 67  
RESULT 46  
AD083966  
ID AD083966 standard; cDNA; 2206 BP.  
XX  
XX AD083966;  
XX  
XX 21-APR-2005 (first entry)  
XX  
XX Plant full length insert polynucleotide seqid 2686.  
DE  
DE plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW protein content; gene; ss.

KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW Galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX Unidentified.  
OS US2004034888-A1.  
PN 19-FEB-2004.  
XX 28-APR-2003; 2003US-00425114.  
XX 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX Claim 1; SEQ ID NO 2686; 15pp; English.  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX Sequence 2206 BP; 691 A; 436 C; 547 G; 532 T; 0 U; 0 Other;  
SQ Query Match 69.6%; Score 17.4; DB 13; Length 2206;  
Best Local Similarity 94.7%; Pred. No. 7.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAAAAGCTGGAATCTCAGGC 20  
|||||  
Db 1215 AAAAAGCTGGAATCTCAGGC 1233  
RESULT 47  
ADX63913  
ID ADX63913 standard; cDNA; 2800 BP.  
XX  
AC ADX63913;  
XX  
DT 21-APR-2005 (first entry)  
XX

DE Plant full length insert polynucleotide seqid 34756.  
XX plant protectant; plant growth regulator; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX Unidentified.  
OS US2004034888-A1.  
PN 19-FEB-2004.  
XX 28-APR-2003; 2003US-00425114.  
XX 06-MAY-1999; 99US-00304517.  
PR 05-NOV-2001; 2001US-00985678.  
XX (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX Claim 1; SEQ ID NO 34756; 15pp; English.  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX Sequence 2800 BP; 835 A; 622 C; 692 G; 651 T; 0 U; 0 Other;  
SQ Query Match 69.6%; Score 17.4; DB 13; Length 2800;  
Best Local Similarity 94.7%; Pred. No. 7.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAAAAGCTGGAATCTCAGGC 20  
|||||  
Db 1801 AAAAAGCTGGAATCTCAGGC 1819  
RESULT 48  
ABA20105/c  
ID ABA20105 standard; DNA; 5161 BP.

XX ABA20105;  
AC  
XX  
XX  
DT 23-JAN-2002 (first entry)  
DE  
XX Human nervous system related polynucleotide SEQ ID NO 12436.  
XX  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
OS  
XX Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
XX 16-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225113P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246521P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-30  
Perfect score: 25  
Sequence: 1 aaaaactggaatctcagctgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

- EST:\*
- 1: gb\_est1:\*
  - 2: gb\_est2:\*
  - 3: gb\_est3:\*
  - 4: gb\_hic:\*
  - 5: gb\_est4:\*
  - 6: gb\_est5:\*
  - 7: gb\_est6:\*
  - 8: gb\_est7:\*
  - 9: gb\_gse1:\*
  - 10: gb\_gse2:\*
  - 11: gb\_gse3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	391	5	BX632940
C 2	25	100.0	435	1	AA963991
C 3	25	100.0	494	2	BE863633
C 4	25	100.0	587	7	CN704489
C 5	25	100.0	771	7	CN461714
C 6	25	100.0	808	2	BE268998
C 7	25	100.0	1717	4	AK043826
C 8	24	96.0	395	5	BY618046
C 9	23.4	93.6	693	10	AG325768
C 10	21.8	87.2	291	2	BB347802
C 11	21.4	85.6	304	2	BB391308
C 12	20.8	83.2	632	6	CA233715
C 13	20.8	83.2	633	6	CA220753
C 14	20.8	83.2	649	6	CA206779
C 15	20.2	80.8	1069	2	BG572464
C 16	20	80.0	362	2	BF589034
C 17	20	80.0	460	2	EG013717
C 18	20	80.0	519	1	AI479209
C 19	20	80.0	581	3	BP250892
C 20	20	80.0	767	5	BX951384
C 21	19.8	79.2	661	9	AQ378916
C 22	19.8	79.2	685	3	BP738553

CO932221	AGENCOURT
AV169292	AV169292
BP070142	BP070142
CA078973	SCRLAM101
BP058178	BP058178
CR868598	Sub scrof
CA253314	SCFLFLA09
DT106989	JGI_ANNR8
BY721049	BY721049
DT107227	JGI_ANNR8
DT105042	JGI_ANNR7
DT110690	JGI_ANNOS
DT118130	JGI_ANNOS
DT100080	JGI_ANNM4
DT104214	JGI_ANNM6
DT105574	JGI_ANNM7
DT108879	JGI_ANNM1
DT093604	JGI_ANNM7
DT072228	JGI_ANNM8
DT095581	JGI_ANNM1
DT105043	JGI_ANNM7
DT103708	JGI_ANNM6
BX241295	DanLo ter
BZ452080	BONPD84TF
BUI25751	603149763
CD126154	MF1-0002U
BU254969	603414046
BU338343	603515585
BU399935	604141211
BU330817	603501552
BU421581	603960428
BU915814	AGENCOURT
AA309289	EST180148
AA309290	EST180149
DE106228	Oryzias 1
AG066899	Pan trogl
CC548878	CH240_433
CK402216	AUF ifint
N85514	J3824F Huma
BF840276	RC4-HT098
BH813513	SALK_0643
CR383055	CR383055
CR383208	CR383208
CR383056	CR383056
CR383209	CR383209
CW739678	MARC_8542
CW739765	MARC_8543
BF955855	RC6-NN023
BB813548	BB813548
BP666638	BP666638
BM481225	533130 MA
AV646888	AV646888
BM433507	10m16002
EX327954	EX327954
AQ015992	CIT-HSP-2
BY450903	BY450903
AV593223	AV593223
CA296670	SCUTSD108
CR382602	CR382602
CF930960	CF--05-R-
BF429789	255676 MA
AV592346	AV592346
AQ420344	RPCI-11-1
RF777309	RF777309
CF930688	CF--04-R-
BF261255	1M0004A10
BG223609	1M0009E06
EG222813	1M00019G0
BG224283	1M0032D09
BF651849	274736 MA
BN967766	LM24HMO11
CR382726	CR382726
AZ259107	RPCI-23-1

77.6	19.4	898	7	CO932221
76.8	19.2	317	1	AV169292
76.8	19.2	410	3	BP070142
76.8	19.2	535	6	CA078973
76.8	19.2	562	3	BP058178
76.8	19.2	637	11	CR868598
76.8	19.2	655	6	CA253314
76.8	19.2	664	8	DT106989
76.8	19.2	672	5	BY721049
76.8	19.2	682	8	DT107227
76.8	19.2	693	8	DT105042
76.8	19.2	694	8	DT110690
76.8	19.2	705	8	DT118130
76.8	19.2	706	8	DT100080
76.8	19.2	712	8	DT104214
76.8	19.2	738	8	DT105574
76.8	19.2	743	8	DT108879
76.8	19.2	755	8	DT093604
76.8	19.2	755	8	DT072228
76.8	19.2	769	8	DT095581
76.8	19.2	772	8	DT105043
76.8	19.2	791	8	DT103708
76.8	19.2	791	10	BX241295
76.8	19.2	824	9	BZ452080
76.8	19.2	861	5	BUI25751
76.0	19	508	6	CD126154
76.0	19	634	5	BU254969
76.0	19	659	5	BU338343
76.0	19	745	5	BU399935
76.0	19	800	5	BU330817
76.0	19	805	5	BU421581
75.2	18.8	950	5	BU915814
75.2	18.8	336	1	AA309289
75.2	18.8	336	1	AA309290
75.2	18.8	613	11	DE106228
75.2	18.8	659	10	AG066899
75.2	18.8	706	9	CC548878
75.2	18.8	892	7	CK402216
74.4	18.6	109	8	N85514
74.4	18.6	266	2	BF840276
74.4	18.6	277	9	BH813513
74.4	18.6	310	7	CR383055
74.4	18.6	310	7	CR383208
74.4	18.6	313	7	CR383056
74.4	18.6	313	7	CR383209
74.4	18.6	331	10	CW739678
74.4	18.6	338	10	CW739765
74.4	18.6	349	2	BF955855
74.4	18.6	380	2	BB813548
74.4	18.6	392	3	BP666638
74.4	18.6	405	3	BM481225
74.4	18.6	436	1	AV646888
74.4	18.6	439	3	BM433507
74.4	18.6	447	5	EX327954
74.4	18.6	452	9	AQ015992
74.4	18.6	453	5	BY450903
74.4	18.6	457	1	AV593223
74.4	18.6	457	6	CA296670
74.4	18.6	458	7	CR382602
74.4	18.6	469	7	CF930960
74.4	18.6	472	2	BF429789
74.4	18.6	478	1	AV592346
74.4	18.6	478	9	AQ420344
74.4	18.6	480	6	CF930688
74.4	18.6	487	7	CF930688
74.4	18.6	490	2	BF261255
74.4	18.6	490	2	BG223609
74.4	18.6	490	2	BG222813
74.4	18.6	494	2	BG224283
74.4	18.6	497	2	BF651849
74.4	18.6	497	3	BN967766
74.4	18.6	502	7	CR382726
74.4	18.6	510	9	AZ259107

96	18.6	74.4	511	2	BG223818	1M00019G1
97	18.6	74.4	511	6	CD178749	MS1-0014P
98	18.6	74.4	516	2	BG224291	IM0032F04
99	18.6	74.4	527	1	AL042149	DKFp434D
100	18.6	74.4	527	8	DT067579	AGENCOURT
101	18.6	74.4	539	1	AW655147	MA
102	18.6	74.4	550	2	BG223824	1M00019H0
103	18.6	74.4	552	7	CN731350	24RDBNH U
104	18.6	74.4	553	3	BM967882	LM24HW009
105	18.6	74.4	554	3	BI898070	478056 MA
106	18.6	74.4	557	1	AW655371	106101 MA
107	18.6	74.4	561	8	DR697427	BP240034B
108	18.6	74.4	564	3	BI847486	469332 MA
109	18.6	74.4	564	9	AZ883187	RPCI-23-2
110	18.6	74.4	565	3	BM967470	LM24HW015
111	18.6	74.4	566	3	BI775377	467824 MA
112	18.6	74.4	568	3	BI774857	467095 MA
113	18.6	74.4	568	3	BM967934	LM24HW008
114	18.6	74.4	573	2	BG224031	1M0026e12
115	18.6	74.4	577	7	CR382803	CE382803
116	18.6	74.4	579	2	BF430176	256134 MA
117	18.6	74.4	584	3	BM967904	LM24HW009
118	18.6	74.4	585	2	BG223923	1M0022F07
119	18.6	74.4	586	3	BM967976	LM24HW008
120	18.6	74.4	592	10	CW739679	MARC 8542
121	18.6	74.4	598	9	AZ104310	RPCI-23-4
122	18.6	74.4	604	3	BM735142	MONOL16
123	18.6	74.4	604	7	CO892850	BovGen_21
124	18.6	74.4	610	6	CF614930	CE5008432
125	18.6	74.4	630	6	CF615053	CE5008334
126	18.6	74.4	643	10	AG014802	Homo sapi
127	18.6	74.4	645	9	AZ266496	RPCI-23-1
128	18.6	74.4	650	6	CB167528	XOV603000
129	18.6	74.4	650	10	AG115741	AG115741 Pan trogl
130	18.6	74.4	652	9	CE199354	tigr-g8s-
131	18.6	74.4	669	5	CF615008	CE5008411
132	18.6	74.4	670	5	BQ849630	QGB10G14
133	18.6	74.4	672	6	CF614496	CE5008673
134	18.6	74.4	674	6	CB463325	723699 MA
135	18.6	74.4	674	7	CO883801	BovGen_12
136	18.6	74.4	675	6	CF615005	CE5008397
137	18.6	74.4	677	7	CN794116	4129246 B
138	18.6	74.4	683	9	CE026656	tigr-g8s-
139	18.6	74.4	689	7	CK962812	4076989 B
140	18.6	74.4	689	7	CR451623	CR451623
141	18.6	74.4	689	7	CR453957	CR453957
142	18.6	74.4	708	9	AZ839429	2M0135K02
143	18.6	74.4	709	7	CK969279	4084687 B
144	18.6	74.4	712	7	CN440684	BE04023A1
145	18.6	74.4	713	5	EX507969	DKFZp6860
146	18.6	74.4	715	5	BY761111	BY761111
147	18.6	74.4	718	1	AM023755	AM023755
148	18.6	74.4	718	7	CK951004	4090406 B
149	18.6	74.4	723	1	AM024344	AM024344
150	18.6	74.4	725	6	CB169006	IMU602702

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 391)  
Henrich, J., Hermanns, J., Kranz, H., Loebbert, R., Schluter, T.,  
Schuette, B., Weindel, M., Heil, O., Neubert, P., Peters, M.,  
Radetot, U., Schneider, D. and Korn, B.  
Mouse ArrayTAG cDNA (LION)  
Unpublished (2003)  
Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; LiONp462C08384.  
RZPDLTB;  
Mouse ArrayTAG cDNA (LION)  
http://www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=4  
62 Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD (clonesrzpd.de) for further information. Seq primer:  
RP: CAGGAACACGTAAGAC.  
Location/Qualifiers  
1 .391  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="LIONp462C08384"  
/lab\_host="DH10B"  
/clone\_lib="pBluescript Lion"

ORIGIN  
Query Match 100.0%; Score 25; DB 5; Length 391;  
Best Local Similarity 100.0%; Pred.No. 1.7;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAACTGGATCTCAGCTGAGA 25  
|||||  
Db 269 AAAAACTGGATCTCAGCTGAGA 245  
|||||

RESULT 2  
AA963991/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA963991  
UI-R-CO-gs-g-01-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone  
UI-R-CO-gs-g-01-0-UI 3', mRNA sequence.  
AA963991  
AA963991.1 GI:4278867  
EST.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
1 (bases 1 to 435)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
8889548  
On May 18, 1998 this sequence version replaced gi:3137483.  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
The sequence tag present in the cDNA between the NorI site and the  
oligo-dr track served to identify it as a clone from the normalized  
adult ovary library. cDNA library preparation: M. Fatima Bonaldo,  
Ph.D. Clone distribution: clones will be available through Research  
Genetics This clone is also available through the I.M.A.G.E.

ALIGNMENTS

RESULT 1  
BX632940/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BX632940  
BX632940 pBluescript Lion 391 bp mRNA linear EST 12-AUG-2003  
3', mRNA sequence.  
BX632940  
BX632940.1 GI:33612815  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1780854

Seq primer: M13 Forward

POLYA=No.

#### FEATURES

source

1. .435  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-C0-98-9-01-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-C0"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-C0  
library is a subtracted library derived from the UI-R-A1  
and UI-R-E1 libraries. The UI-R-A1 library consisted of a  
mixture of individually tagged normalized libraries  
constructed from rat placenta, adult lung, brain, liver,  
kidney, heart, spleen, ovary, and muscle. The UI-R-E1  
library consisted of a mixture of individually tagged  
normalized libraries constructed from 8, 12 and 18-day  
embryo. The tag is a string of 3-5 nucleotides present  
between the Not I site and the oligo-dT track which  
allows identification of the library of origin of a clone  
within the mixture. The subtracted library (UI-R-C0) was  
constructed as follows: PCR amplified cDNA inserts from a  
pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had  
been derived was used as a driver in a hybridization with  
the pooled UI-R-A1 and UI-R-E1 library in the form of  
single-stranded circles. The remaining single-stranded  
circles (subtracted library) was purified by  
hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-C0  
library. This procedure has been previously described  
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,  
1996)"

#### FEATURES

source

1. 494  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH0-ake-c-06-0-UI"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BMAP\_M\_S1"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The  
NIH BMAP\_M\_S1 library is a subtracted library derived from  
a mixture of normalized libraries from ten regions of the  
mouse brain (cerebellum, brain stems, olfactory bulbs,  
hypothalamus, cortex, amygdala, basal ganglia, pineal  
gland, striatum, hippocampus). The driver used for  
subtraction consisted of a pool of 20,000 cDNA clones  
obtained from non-normalized and normalized libraries of  
these ten regions of the mouse brain."

#### ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 494;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25

Db 296 AAAAACTGGAATCTCAGGCTGAGA 272

#### ORIGIN

Query Match 100.0%; Score 25; DB 1; Length 435;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25

Db 314 AAAAACTGGAATCTCAGGCTGAGA 290

#### RESULT 3

BE863633/c

LOCUS

DEFINITION UI-M-BH0-ake-c-06-0-UI.r1 NIH BMAP\_M\_S1 Mus musculus cDNA clone  
494 bp mRNA linear EST 29-SEP-2000

UI-M-BH0-ake-c-06-0-UI 5', mRNA sequence.

BE863633

BE863633.1 GI:10383871

EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 494)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
should be noted that Bento Soares is generating a small number of  
additional specialized non-redundant arrays of BMAP cDNAs whose  
availability will be considered under appropriate and limited  
collaborative arrangements

Seq primer: M13 Reverse.

Location/Qualifiers

1. 494

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH0-ake-c-06-0-UI"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NIH BMAP\_M\_S1"

/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The  
NIH BMAP\_M\_S1 library is a subtracted library derived from  
a mixture of normalized libraries from ten regions of the  
mouse brain (cerebellum, brain stems, olfactory bulbs,  
hypothalamus, cortex, amygdala, basal ganglia, pineal  
gland, striatum, hippocampus). The driver used for  
subtraction consisted of a pool of 20,000 cDNA clones  
obtained from non-normalized and normalized libraries of  
these ten regions of the mouse brain."

#### RESULT 4

CN704489

LOCUS

DEFINITION

musculus cDNA clone NIA:E0488D11 IMAGE:30877006 5', mRNA sequence.

CN704489

EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 587)

VanBuren, V., Falco, G., Martin, P.R., Stagg, C.A., Bassey, U.C.,

Wang, Y., Carter, M.G., Hamatani, T., Aiba, K., Akutsu, H., Sharova, L.,

Nagajima, R., Boheler, K.R., Taub, D., Hodes, R.J., Longo, D.L.,

Schlesinger, D., Keller, J., Klotz, E., Kelsae, G., Umezawa, A.,

Vescovi, A.L., Roseant, J., Kunath, T., Hogan, B.L., Curci, A.,

D'Urso, M., Kelsae, J., Hide, W. and Ko, M.S.

Transcriptome analysis of mouse stem cells and early embryos

PLoS Biol. 1 (3), 410-419 (2003)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: E0488 row: D column: 11

Seq primer: M13 Reverse

High quality sequence stop: 587

POLYA=No.

FEATURES  
source

Location/Qualifiers  
1. .587  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:E0488D11-5"  
/db\_xref="taxon:10090"  
/clone="NIA:E0488D11 IMAGE:30877006"  
/tissue\_type="whole embryo including extraembryonic  
tissues at 11.5-days postcoitum"  
/dev\_stage="E11.5"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse E11.5 whole embryo cDNA library  
(Long)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>).  
This is a long-transcript enriched cDNA library (Ref.  
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total  
RNAs were extracted from a pool of 3 embryos at 11.5-days  
postcoitum. Double-stranded cDNAs were synthesized with an  
Oligo(dT) primer [Invitrogen:  
5'-pGACTAGTCTAGATCGGAGCGCCGCTTTT-3'] from  
2ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to Loner-linker L1-Sal4, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricion  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 3.3Kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 587;

Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25

Db 121 AAAAACTGGAATCTCAGGCTGAGA 145

## RESULT 5

CN461714

LOCUS

DEFINITION CN461714 771 bp mRNA linear EST 21-APR-2004  
IMAGE:30650666 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>

This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5

FEATURES  
source

Location/Qualifiers  
1. .771  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30650666"  
/tissue\_type="Upper Head"  
/dev\_stage="9.5-10.5 dpc"  
/lab\_host="DH10B (T1 phage  
resistant)"  
/clone\_lib="NIH BMAP HN0"

/note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CCAACTGAAT. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 771;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25

Db 90 AAAAACTGGAATCTCAGGCTGAGA 114

## RESULT 6

BB268998

LOCUS

DEFINITION BB268998 RIKEN full-length enriched, 10 days neonate cortex Mus  
musculus cDNA clone A830037L14 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 808)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,

Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

On Jul 7, 2000 this sequence version replaced gi:8965468.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.jp](mailto:genome-res@gsc.riken.jp), <http://genome.gsc.riken.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.





```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A830037L14"
/db_xref="taxon:10090"
/clone="A830037L14"
/tissue_type="cortex"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
misc_feature
1..1717
/note="unknown EST (GB|BB641983, evidence: BLASTN, 100%,
match=519)"

ORIGIN
Query Match 100.0%; Score 25; DB 4; Length 1717;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 1456 AAAAACTGGAATCTCAGGCTGAGA 1480

RESULT 8
BY618046 395 bp mRNA linear EST 15-DEC-2002
LOCUS BY618046 RIKEN full-length enriched, visual cortex Mus musculus
DEFINITION cDNA clone K330032L03 3', mRNA sequence.
ACCESSION BY618046
VERSION BY618046.1 GI:26953228
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 395)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Mutsaers,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L.G., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

```

Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
 Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
 Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
 Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
 Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
 Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Michela Fagioli and Takao K. Hensch (  
 Laboratory for Neuronal Circuit Development Brain Science Institute  
 RIKEN 2-1 Hirotsawa,Wako-shi,Saitama 351-0198 Japan ) whose  
 assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.

FEATURES                      Location/Qualifiers  
                                  1..395  
                                  /organism="Mus musculus"  
                                  /mol\_type="mRNA"  
                                  /strain="C57BL/6J"  
                                  /db\_xref="taxon:10090"  
                                  /clone="K330032L03"  
                                  /tissue\_type="visual cortex"  
                                  /clone\_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 96.0%; Score 24; DB 5; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
      |||||  
 Db 114 AAAAACTGGAATCTCAGGCTGAGA 137

RESULT 9  
 AG325768 Mus musculus molossinus DNA, clone:MSMg01-113L04.T7, genomic survey  
 LOCUS AG325768 sequence.  
 DEFINITION AG325768 Mus musculus molossinus (Japanese wild mouse)  
 ACCESSION AG325768  
 VERSION AG325768.1 GI:47898759  
 KEYWORDS GSS.  
 SOURCE Mus musculus molossinus  
 ORGANISM Mus musculus molossinus (Japanese wild mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,  
 Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and  
 Shiroishi,T.  
 Contribution of Asian mouse subspecies Mus musculus molossinus to  
 genomic constitution of strain C57BL/6J, as defined by BAC-end  
 sequence-SNP analysis  
 Genome Res. 14 (12), 2439-2447 (2004)  
 PUBMED 15574823





```

ACCESSION      CA220753
VERSION        CA220753.1  GI:35274377
KEYWORDS       EST.
SOURCE         Saccharum officinarum
ORGANISM       Saccharum officinarum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
               clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
               complex.
REFERENCE      1 (bases 1 to 633)
AUTHORS       Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE         The libraries that made SUCEST
JOURNAL       Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT       Contact: Arruda P
               Centro de Biologia Molecular e Engenharia Genetica
               Universidade Estadual de Campinas
               Caixa Postal 6010, 13083-970, Campinas SP, Brazil
               Tel: 55 19 3788 1137
               Fax: 55 19 3788 1089
               Email: parruda@unicamp.br
               Clone distribution: clone distribution information can be found
               through the Brazilian Clone Collection Center (BCCC) at
               http://www.bcccenter.fcav.unesp.br
               Plate: 124 row: A column: 01
               Seq primer: T7 Promoter Primer.
               Location/Qualifiers
               1..649
                /organism="Saccharum officinarum"
                /mol_type="mRNA"
                /db_xref="taxon:4547"
                /clone="SCACSB1124A01"
                /lab_host="DH10B"
                /clone_lib="SBI"
                /note="Organ: Stalk Bark from adult plants; Vector:
                pSport1; Site 1: SalI; Site 2: NotI; An unidirectional
                cDNA library generated from [Stalk Bark from adult
                plants]. cDNA was prepared from polyA+ mRNA using
                SuperScript Plasmid System Kit (Invitrogen). The
                double-strand cDNAs were fractionated in a sepharose
                CL-2B 40cm-columns and fragments sizing between 0.8 and
                1.5 Kb were directionally cloned into the vector. Details
                of each source of RNA and library construction can be
                obtained at http://sucst.lad.ic.unicamp.br/public"

FEATURES             source
    source
ORIGIN
Query Match      83.2%; Score 20.8; DB 6; Length 649;
Best Local Similarity 91.7%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2  AAAAAGCTGGAATCTCAGGCTGAGA 25
    ||||| ||||| ||||| ||||| |||||
Db  60  AAAAAGCTGGAATCTCAGGCTGAAA 83

RESULT 15
BG572464/c      1069 bp      mRNA      linear      EST 10-APR-2001
LOCUS           602593518F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4720829 5',
DEFINITION      mRNA sequence.
ACCESSION       BG572464
VERSION         BG572464.1  GI:13580117
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Hominiidae; Homo.
REFERENCE      1 (bases 1 to 1069)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: CLONTECH Laboratories, Inc.
               cDNA Library Preparation: CLONTECH Laboratories, Inc.
               DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCMI576 row: c column: 06
               High quality sequence stop: 448.
               Location/Qualifiers
               1..1069
                /organism="Homo sapiens"
                /mol_type="mRNA"

FEATURES             source
    source

```

```
/db_xref="taxon:9606"
/clone="IMAGE:4720829"
/lab_host="DH108 (T1 phage-resistant)"
/clone_lib="NIH_MGC_79"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
(ggcctatggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGCGCCGACATG-dT(30)EN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.3
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
```

## ORIGIN

```
Query Match      80.8%; Score 20.2; DB 2; Length 1069;
Best Local Similarity 88.0%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
||| ||||| ||||| ||||| |||||
Db 765 AAACRACTGGAATGTCAGGCTGAAA 741
```

## RESULT 16

```
BF589034/c
LOCUS      362 bp mRNA linear EST 12-DEC-2000
DEFINITION naa40f10.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:3258787
3' mRNA sequence.
```

```
ACCESSION   BF589034
VERSION     BF589034.1 GI:11681358
KEYWORDS    EST.
```

## SOURCE

```
ORGANISM    Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 362)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
AUTHORS     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

## TITLE

```
Tumor Gene Index
Unpublished (1997)
```

```
Contact: Robert Strausberg, Ph.D.
```

```
Email: cgaps-r@mail.nih.gov
```

```
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
```

```
Emmert-Buck, M.D., Ph.D.
```

```
CDNA Library Preparation: M. Bento Soares, Ph.D.
```

```
CDNA Library Arrayed by: Greg Lennon, Ph.D.
```

```
DNA Sequencing by: Washington University Genome Sequencing Center
```

```
Clone Distribution: NCI-CGAP clone distribution information can be
```

```
found through the I.M.A.G.E. Consortium/LINL, send email to:
```

```
info@image.llnl.gov
```

```
Seq primer: -40UP from Gibco
```

```
High quality sequence stop: 294.
```

## FEATURES

```
source
1..362
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3258787"
/lab_host="DH108"
/clone_lib="NCI CGAP Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
```

## ORIGIN

```
Query Match      80.0%; Score 20; DB 2; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 AAACCTGGAATCTCAGGCTGA 23
||||| ||||| ||||| ||||| |||||
```

```
Db 272 AAACCTGGAATCTCAGGCTGA 253
```

## RESULT 17

```
BF589034/c
LOCUS      460 bp mRNA linear EST 24-JAN-2001
DEFINITION MRI-GN0173-181200-023-b07 GN0173 Homo sapiens cDNA, mRNA sequence.
```

## ACCESSION

```
BF589034
```

## VERSION

```
BF589034.1 GI:12464186
```

## KEYWORDS

```
EST.
```

## SOURCE

```
ORGANISM    Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
```

## REFERENCE

```
1 (bases 1 to 460)
```

## AUTHORS

```
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
```

```
Shotgun sequencing of the human transcriptome with ORF expressed
```

## sequence tags

```
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
```

## JOURNAL

```
PUBMED
```

## COMMENT

```
Contact: Simpson A.J.G.
```

```
Laboratory of Cancer Genetics
```

```
Ludwig Institute for Cancer Research
```

```
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
```

```
Brazil
```

```
Tel: +55-11-2704922
```

```
Fax: +55-11-2707001
```

```
Email: asimpson@ludwig.org.br
```

```
This sequence was derived from the FAPESP/LICR Human Cancer Genome
```

```
Project. This entry can be seen in the following URL
```

```
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR1&t2=MR1-GN0173-
```

```
181200-023-b07&t3=2000-12-18&t4=1)
```

```
Seq primer: puc 18 forward
```

```
High quality sequence start: 20
```

```
High quality sequence stop: 460.
```

## FEATURES

## source

```
1..460
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="GN0173"
/note="Organ: placenta normal; Vector: puc18; Site 1:
SmaI; Site 2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
```

## ORIGIN

```
Query Match      80.0%; Score 20; DB 2; Length 460;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 AAACCTGGAATCTCAGGCTGA 23
||||| ||||| ||||| ||||| |||||
Db 204 AAACCTGGAATCTCAGGCTGA 223
```

```

RESULT 18
AI479209/c
LOCUS
DEFINITION
tm55904.x1 NCI_CGAP_kid11 Homo sapiens cDNA clone IMAGE:2162070 3',
mRNA sequence.
ACCESSION
AI479209
VERSION
AI479209.1 GI:4372377
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 519)
NCI-CGAP http://www.ncbi.nlm.nih.gov/nciogap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strauberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 720 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 465.
FEATURES
source
1..519
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2162070"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Kid11"
/notes="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid1 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneId8 1323376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
ORIGIN
Query Match 80.0%; Score 20; DB 1; Length 519;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAACCTGGAATCTCAGGCTGA 23
|||||
Db 271 AAACCTGGAATCTCAGGCTGA 252

RESULT 19
BP250892
LOCUS
DEFINITION
BP250892 Sugano cDNA library, hippocampus Homo sapiens cDNA clone
HPR05485, mRNA sequence.
ACCESSION
BP250892
VERSION
BP250892.1 GI:52133171
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 581)
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HPR05485"
/tissue_type="hippocampus"
/clone_lib="Sugano cDNA library, hippocampus"
ORIGIN
Query Match 80.0%; Score 20; DB 3; Length 581;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAACCTGGAATCTCAGGCTGA 23
|||||
Db 94 AAACCTGGAATCTCAGGCTGA 113

RESULT 20
BX951384
LOCUS
DEFINITION
DKFZP781E09126_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
DKFZP781E09126 5', mRNA sequence.
ACCESSION
BX951384
VERSION
BX951384.1 GI:43426263
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 767)
Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
Unpublished (2003)
Contact: MIPS
JOURNAL
COMMENT
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZP781E09126) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZP781E09126"
/dev_stage="adult"

```

```

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 581)
AUTHORS
Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL
Genome Res. 14 (9), 1711-1718 (2004)
PUBMED
15342556
COMMENT
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.
FEATURES
Location/Qualifiers
1..581
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HPR05485"
/tissue_type="hippocampus"
/clone_lib="Sugano cDNA library, hippocampus"
ORIGIN
Query Match 80.0%; Score 20; DB 3; Length 581;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAACCTGGAATCTCAGGCTGA 23
|||||
Db 94 AAACCTGGAATCTCAGGCTGA 113

```



```

/lab_host="DH10B"
/clone_lib="781 (synonym: hlcc4)"
/notes="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN
Query Match      80.0%; Score 20; DB 5; Length 767;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGA 23
    |||||
Db 708 AAACCTGGAATCTCAGGCTGA 727

RESULT 21
AQ378916
LOCUS
DEFINITION
  AQ378916 661 bp DNA linear GSS 20-MAY-1999
  genomic survey sequence.
ACCESSION
  AQ378916
VERSION
  AQ378916.1 GI:4349939
KEYWORDS
  GSS.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
    Homnidae; Homo.
REFERENCE
  Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
  Venter,J.C
  1 (bases 1 to 661)
  Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
  Map Building
  Unpublished (1997)
JOURNAL
  Other GSSs: RPCI11-164114.TV
COMMENT
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@jcm.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genetics (info@resgen.com). BAC end search page:
  http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1..661
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="GDB:7562797"
        /db_xref="taxon:9606"
        /clone="RPCI-11-164114"
        /sex="Male"
        /cell_type="Lymphocytes"
        /clone_lib="RPCI-11"
        /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
        RPCI11 Human Male BAC Library"

ORIGIN
Query Match      79.2%; Score 19.8; DB 9; Length 661;
Best Local Similarity 91.3%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAG 24
    |||||
Db 550 AGAATCTGGAATCTCAGGCTGAG 572

```

```

RESULT 22
BP738553/c
LOCUS
DEFINITION
  BP738553 Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA
  library Xenopus laevis cDNA clone XL516n04ex 3', mRNA sequence.
ACCESSION
  BP738553
VERSION
  BP738553.1 GI:46087146
KEYWORDS
  EST.
SOURCE
  Xenopus laevis (African clawed frog)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
    Xenopodinae; Xenopus; Xenopus.
  1 (bases 1 to 685)
  Osada,S., Kitayama,A., Ueno,N. and Taira,M.
  Expression analysis of genes which are expressed in the anterior
  neuroectoderm of Xenopus embryos
  Unpublished (2004)
JOURNAL
  Contact: Masanori Taira
  Department of Biological Sciences
  Graduate School of Science, University of Tokyo; CREST, Japan
  Science and Technology Corporation, Japan
  7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
  Tel: 81-03-5841-4434
  Fax: 81-03-5841-4434
  Email: m.taira@biol.s.u-tokyo.ac.jp,
  URL:http://www.shigen.nig.ac.jp/nbrp/xenopus/est/.
  Location/Qualifiers
    1..685
      /organism="Xenopus laevis"
      /mol_type="mRNA"
      /db_xref="taxon:8355"
      /clone="XL516n04ex"
      /tissue_type="anterior neuroectoderm"
      /dev_stage="late gastrula (stage 12.5)"
      /clone_lib="Osada Taira anterior neuroectoderm (ANE)
      PCS105 cDNA library"

ORIGIN
Query Match      79.2%; Score 19.8; DB 3; Length 685;
Best Local Similarity 91.3%; Pred. No. 5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGA 23
    |||||
Db 492 AAAAAACGAGGAATCTCAGGATGA 470

RESULT 23
CO932221
LOCUS
DEFINITION
  CO932221 898 bp mRNA linear EST 16-AUG-2004
  AGENCOURT_30436848 NIH_ZGC_14 Danio rerio cDNA clone IMAGE:7406857
  5', mRNA sequence.
ACCESSION
  CO932221
VERSION
  CO932221.1 GI:51286898
KEYWORDS
  EST.
SOURCE
  Danio rerio (zebrafish)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
  1 (bases 1 to 898)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Daniela S. Gerhard, Ph.D.
  Office of Cancer Genomics / NIH
  National Cancer Institute
  Bldg. 31 Rm10A07 Bethesda, MD 20892
  Email: cgabbs-r@mail.nih.gov
  Tissue Procurement: John Ngai, Nancy Freeman, NIDCD
  cDNA Library Preparation: Dr. Sumio Sugano
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```





```

RESULT 26
CA078973
LOCUS       SCRLAM1013F07.1 535 bp mRNA linear EST 23-SEP-2003
DEFINITION   SCRLAM1013F07.1 Saccharum officinarum cDNA clone SCRLAM1013F07
5', mRNA sequence.
ACCESSION   CA078973
VERSION     CA078973.1 GI:34931245
KEYWORDS    EST.
SOURCE      Saccharum officinarum
ORGANISM    Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1 (bases 1 to 535)
AUTHORS    Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE      The libraries that made SUCBST
JOURNAL    Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT    Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 013 row: F column: 07
            Seq primer: T7 Promoter Primer.
            Location/Qualifiers
                1..535
                /organism="Saccharum officinarum"
                /mol_type="mRNA"
                /db_xref="taxon:4547"
                /clone="SCRLAM1013F07"
                /lab_host="DH10B"
                /clone_lib="AM1"
                /notes="Organ: Apical meristem and tissues surrounding of
                mature plants; Vector: pSport1; Site 1: SalI; Site 2:
                NotI; An unidirectional cDNA library generated from
                [apical meristem and tissues surrounding of mature
                plant]. cDNA was prepared from polyA+ mRNA using
                SuperScript Plasmid System Kit (Invitrogen). The
                double-strand cDNAs were fractionated in a sepharose
                CL-2B 40cm-columns and fragments sizing between 0.8 and
                1.5 Kb were directionally cloned into the vector. Details
                of each source of RNA and library construction can be
                obtained at http://sucst.lad.ic.unicamp.br/public"

FEATURES             source
    source
    1..535
    /organism="Saccharum officinarum"
    /mol_type="mRNA"
    /db_xref="taxon:4547"
    /clone="SCRLAM1013F07"
    /lab_host="DH10B"
    /clone_lib="AM1"
    /notes="Organ: Apical meristem and tissues surrounding of
    mature plants; Vector: pSport1; Site 1: SalI; Site 2:
    NotI; An unidirectional cDNA library generated from
    [apical meristem and tissues surrounding of mature
    plant]. cDNA was prepared from polyA+ mRNA using
    SuperScript Plasmid System Kit (Invitrogen). The
    double-strand cDNAs were fractionated in a sepharose
    CL-2B 40cm-columns and fragments sizing between 0.8 and
    1.5 Kb were directionally cloned into the vector. Details
    of each source of RNA and library construction can be
    obtained at http://sucst.lad.ic.unicamp.br/public"

ORIGIN
Query Match       76.8%; Score 19.2; DB 6; Length 535;
Best Local Similarity 87.5%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCGGAATCTCAGGCTGAGA 25
|||||
Db 376 AAAAGCTGGAATCTCAAGCTGAAA 399
|||||

RESULT 27
BP058178
LOCUS       BP058178.1 562 bp mRNA linear EST 25-AUG-2004
DEFINITION   BP058178 Lotus corniculatus var. japonicus pods (less than 20 mm in
length) Lotus corniculatus var. japonicus cDNA clone SPDL098f12_f
3', mRNA sequence.
ACCESSION   BP058178
VERSION     BP058178.1 GI:45614929
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM    Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
            Lotus.
            1 (bases 1 to 562)
            Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
            Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
            Plant Mol. Biol. 54 (3), 405-414 (2004)
            15284495
            Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
            Location/Qualifiers
                1..562
                /organism="Lotus corniculatus var. japonicus"
                /mol_type="mRNA"
                /isolate="Miyakojima MG-20"
                /db_xref="taxon:34305"
                /clone="SPDL098f12_f"
                /tissue_type="pods"
                /dev_stages="pod (less than 20 mm in length)"
                /clone_lib="Lotus corniculatus var. japonicus pods (less
                than 20 mm in length)"

ORIGIN
Query Match       76.8%; Score 19.2; DB 3; Length 562;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCGGAATCTCAGGCTGAGA 25
|||||
Db 153 AAAAATCGGAATCTCAGGCAAGA 176
|||||

RESULT 28
CR868598/c
LOCUS       CR868598 637 bp DNA linear GSS 19-NOV-2004
DEFINITION   Sus scrofa BES, genomic survey sequence.
ACCESSION   CR868598
VERSION     CR868598.1 GI:55866844
KEYWORDS    GSS; Bac-end sequence BES; Genome Survey Sequence.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
            1 (bases 1 to 637)
            Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
            Chardon,P.
            Construction of a swine BAC library: application to the
            characterization and mapping of porcine type C endoviral elements
            Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)
            10449899
            2 (bases 1 to 637)
            Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J.,
            Rogel-Gaillard,C., Roy,A., Schibler,B. and Milan,D.
            A physical map of the swine genome
            Unpublished
            3 (bases 1 to 637)
            Genoscope.
            Direct Submission
            Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr
            Location/Qualifiers
                1..637
                /organism="Sus scrofa"
                /mol_type="genomic DNA"
                /strain="Large White"
                /db_xref="taxon:9823"

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

1 (bases 1 to 562)  
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
Characteristics of the Lotus japonicus gene repertoire deduced from large-scale expressed sequence tag (EST) analysis  
Plant Mol. Biol. 54 (3), 405-414 (2004)  
15284495  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers

1..562

/organism="Lotus corniculatus var. japonicus"  
/mol\_type="mRNA"  
/isolate="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="SPDL098f12\_f"  
/tissue\_type="pods"  
/dev\_stages="pod (less than 20 mm in length)"  
/clone\_lib="Lotus corniculatus var. japonicus pods (less than 20 mm in length)"

ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 562;  
Best Local Similarity 87.5%; Pred. No. 9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAATCGGAATCTCAGGCTGAGA 25

Db 153 AAAAATCGGAATCTCAGGCAAGA 176

RESULT 28

CR868598/c  
LOCUS CR868598 637 bp DNA linear GSS 19-NOV-2004  
DEFINITION Sus scrofa BES, genomic survey sequence.  
ACCESSION CR868598  
VERSION CR868598.1 GI:55866844  
KEYWORDS GSS; Bac-end sequence BES; Genome Survey Sequence.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 637)

Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and Chardon,P.

Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)

10449899

2 (bases 1 to 637)

Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J.,

Rogel-Gaillard,C., Roy,A., Schibler,B. and Milan,D.

A physical map of the swine genome

Unpublished

3 (bases 1 to 637)

Genoscope.

Direct Submission

Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

Location/Qualifiers

1..637

/organism="Sus scrofa"

/mol\_type="genomic DNA"

/strain="Large White"

/db\_xref="taxon:9823"

BI/21049/C	LOCUS	BY721049	672 bp	mRNA	linear	EST 17-DEC-2002
	DEFINITION	BY721049	RIKEN	full-length enriched,	adult male	epididymis Mus



the JGI Clone Id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.  
 Poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated.  
 The resulting Poly-T sequence has been removed.  
 Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.  
 Plate: ANNN 0089 row: 4 column: 7  
 High quality sequence stop: 556  
 POLYA=Yes.

#### FEATURES

#### Location/Qualifiers

1. .682

/organism="Pimephales promelas"

/mol\_type="mRNA"

/db\_xref="taxon:90988"

/clone="ANNN8473"

/tissue\_type="Whole"

/dev\_stage="Adult"

/clone\_lib="ANNN Pimephales promelas Whole (L)"

/notes="Vector: pDONR222; The library was made from dT primed cDNA and cloned into Invitrogen vector pDONR222. Poly A RNA were primed with a Biotin-attB2-Oligo(dT) primer (5'- GGCGCGCACAACTTTGTACAGAAAGTTGGGT(T)19 -3') and the first strand synthesized using Superscript II (Invitrogen). The second strand was synthesized and an attB1 adaptor was ligated to the 5' end (5'- TCCTCGGGGCAACTTTGTACAAAGTTGCC -3') . cDNA was size selected using 1% agarose gel electrophoresis (L ~0.5-1.2k, M ~1.2-2.5k, H ~2.5k) and then inserted into the vector using site specific recombination (flanking attB sites on cDNA). The work was done at DOE Joint Genome Institute."

2 AAAAACTGGAATCTCAGGCTGAGA 25

|||||

604 AAAAACTGCAATATCAGCCTGAGA 627

Db

Query Match 76.8%; Score 19.2; DB 8; Length 682;

Best Local Similarity 87.5%; Pred. No. 9.4e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

Db

RESULT 33

DT105042

LOCUS

DEFINITION

JGI ANNN7222.rev ANNN Pimephales promelas Whole (L) Pimephales

promelas cDNA clone ANNN7222 3', mRNA sequence.

DT105042

VERSION

KEYWORDS

SOURCE

ORGANISM

Pimephales promelas

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Pimephales.

1 (bases 1 to 693)

Richardson, P., Lucas, S., Rokhsar, D., Detter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Pimephales promelas EST project

Unpublished (2005)

Other ESTs: JGI ANNN7222.fwd

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

cDNA Library Preparation: DOE Joint Genome Institute:

http://www.jgi.doe.gov

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.  
 Poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated.  
 The resulting Poly-T sequence has been removed.  
 Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.  
 Plate: ANNN 0073 row: 1 column: 6  
 High quality sequence stop: 554  
 POLYA=Yes.

#### FEATURES

#### Location/Qualifiers

1. .693

/organism="Pimephales promelas"

/mol\_type="mRNA"

/db\_xref="taxon:90988"

/clone="ANNN7222"

/tissue\_type="Whole"

/dev\_stage="Adult"

/clone\_lib="ANNN Pimephales promelas Whole (L)"

/notes="Vector: pDONR222; The library was made from dT primed cDNA and cloned into Invitrogen vector pDONR222. Poly A RNA were primed with a Biotin-attB2-Oligo(dT) primer (5'- GGCGCGCACAACTTTGTACAGAAAGTTGGGT(T)19 -3') and the first strand synthesized using Superscript II (Invitrogen). The second strand was synthesized and an attB1 adaptor was ligated to the 5' end (5'- TCCTCGGGGCAACTTTGTACAAAGTTGCC -3') . cDNA was size selected using 1% agarose gel electrophoresis (L ~0.5-1.2k, M ~1.2-2.5k, H ~2.5k) and then inserted into the vector using site specific recombination (flanking attB sites on cDNA). The work was done at DOE Joint Genome Institute."

2 AAAAACTGGAATCTCAGGCTGAGA 25

|||||

604 AAAAACTGCAATATCAGCCTGAGA 627

Db

Query Match 76.8%; Score 19.2; DB 8; Length 693;

Best Local Similarity 87.5%; Pred. No. 9.4e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

Db

RESULT 34

DT110690/c

LOCUS

DEFINITION

JGI ANNO998.fwd ANNO Pimephales promelas Whole (M) Pimephales

promelas cDNA clone ANNO998 5', mRNA sequence.

DT110690

VERSION

KEYWORDS

SOURCE

ORGANISM

Pimephales promelas

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Pimephales.

1 (bases 1 to 694)

Richardson, P., Lucas, S., Rokhsar, D., Detter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Pimephales promelas EST project

Unpublished (2005)

Other ESTs: JGI ANNO998.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

cDNA Library Preparation: DOE Joint Genome Institute:

http://www.jgi.doe.gov

DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
 Naming Conventions: EST name is generated by the concatenation of  
 the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
 indicates a forward sequencing read of the insert. It does not  
 necessarily reflect the orientation of the insert.  
 Small Insert: Based upon one or more sequencing reads of this clone  
 where vector sequence was present at both ends, this clone has been  
 determined to contain a cDNA insert on the order of 600-1000 bases.  
 High quality sequence stop: 681.

#### FEATURES

source

```

1. 694
Location/Qualifiers
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNO998"
/dev_stages="Adult"
/tissue_type="Whole"
/clone_lib="ANNO Pimephales promelas Whole (M)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'- GGCGGCGCACAACTTTGTACAAGAAAGTTGGT(τ)19 -3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
attB1 adaptor was ligated to the 5' end (5'-
TCGTCGGGACAACTTTGTACAAAAGTTGG -3' and 5'-
CCAACTTTTGTACAAAGTTGCCCC -3'). The cDNA was size
selected using 1% agarose gel electrophoresis. (L
-0.5-1.2k, M -1.2-2.5k, H -2.5k) and then inserted into
the vector using site specific recombination (flanking
attB sites on cDNA). The work was done at DOE Joint Genome
Institute."
```

#### ORIGIN

```

Query Match          76.8%; Score 19.2; DB 8; Length 694;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2  AAAAACTGGAATCTCAGGCTGAGA 25
      ||||| ||||| ||||| ||||| |||||
Db   144 AAAAACTGCAATATCAGCCTGAGA 121
```

#### RESULT 35

```

DT118130/c
LOCUS          705 bp      mRNA      linear      EST 12-AUG-2005
DEFINITION    JGI_ANNO5327.fwd ANNO Pimephales promelas Whole (M) Pimephales
              promelas cDNA clone ANNO5327 5', mRNA sequence.
ACCESSION     DT118130
VERSION       DT118130.1
KEYWORDS      GI:72482755
SOURCE        Pimephales promelas
ORGANISM      Pimephales promelas
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Pimephales.
REFERENCE     1 (bases 1 to 705)
              Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
              Brokstein,P. and Lindquist,E.A.
              DOE Joint Genome Institute Pimephales promelas EST project
              Unpublished (2005)
              Other ESTs: JGI ANNO5327.rev
              Contact: Lindquist,E.A., Richardson,P.
              DOE Joint Genome Institute
              2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              Tel: 925 296 5600
              Fax: 925 296 5710
              Email: cdna@jgi-psf.org
              cDNA Library Preparation: DOE Joint Genome Institute:
              http://www.jgi.doe.gov
              Naming Conventions: EST name is generated by the concatenation of
              the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
              indicates a forward sequencing read of the insert. It does not
```

the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
 indicates a forward sequencing read of the insert. It does not  
 necessarily reflect the orientation of the insert.  
 Small Insert: Based upon one or more sequencing reads of this clone  
 where vector sequence was present at both ends, this clone has been  
 determined to contain a cDNA insert on the order of 600-1000 bases.  
 Plate: ANNO 0053 row: n column: 12  
 High quality sequence stop: 624.

#### FEATURES

source

```

1. 705
Location/Qualifiers
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNO5327"
/dev_stages="Adult"
/tissue_type="Whole"
/clone_lib="ANNO Pimephales promelas Whole (M)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'- GGCGGCGCACAACTTTGTACAAGAAAGTTGGT(τ)19 -3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
attB1 adaptor was ligated to the 5' end (5'-
TCGTCGGGACAACTTTGTACAAAAGTTGG -3' and 5'-
CCAACTTTTGTACAAAGTTGCCCC -3'). The cDNA was size
selected using 1% agarose gel electrophoresis. (L
-0.5-1.2k, M -1.2-2.5k, H -2.5k) and then inserted into
the vector using site specific recombination (flanking
attB sites on cDNA). The work was done at DOE Joint Genome
Institute."
```

#### ORIGIN

```

Query Match          76.8%; Score 19.2; DB 8; Length 705;
Best Local Similarity 87.5%; Pred. No. 9.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2  AAAAACTGGAATCTCAGGCTGAGA 25
      ||||| ||||| ||||| ||||| |||||
Db   128 AAAAACTGCAATATCAGCCTGAGA 105
```

#### RESULT 36

```

DT100080/c
LOCUS          706 bp      mRNA      linear      EST 12-AUG-2005
DEFINITION    JGI_ANNA4389.fwd ANNN Pimephales promelas Whole (L) Pimephales
              promelas cDNA clone ANNA4389 5', mRNA sequence.
ACCESSION     DT100080
VERSION       DT100080.1
KEYWORDS      GI:72445574
SOURCE        Pimephales promelas
ORGANISM      Pimephales promelas
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Pimephales.
REFERENCE     1 (bases 1 to 706)
              Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
              Brokstein,P. and Lindquist,E.A.
              DOE Joint Genome Institute Pimephales promelas EST project
              Unpublished (2005)
              Other ESTs: JGI ANNA4389.rev
              Contact: Lindquist,E.A., Richardson,P.
              DOE Joint Genome Institute
              2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              Tel: 925 296 5600
              Fax: 925 296 5710
              Email: cdna@jgi-psf.org
              cDNA Library Preparation: DOE Joint Genome Institute:
              http://www.jgi.doe.gov
              Naming Conventions: EST name is generated by the concatenation of
              the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
              indicates a forward sequencing read of the insert. It does not
```







FEATURES  
source

Location/Qualifiers  
1..738  
/organism="Pimephales promelas"  
/mol\_type="mRNA"  
/db\_xref="taxon:90988"  
/clone="ANN7528"  
/tissue\_type="Whole"  
/dev\_stage="Adult"  
/clone\_lib="ANN Pimephales promelas Whole (L)"  
/note="Vector: pDONR222; The library was made from dT primed cDNA and cloned into Invitrogen vector pDONR222. Poly A RNA were primed with a Biotin-attB2-Oligo(dT) primer (5'-GGCGCGCACAACTTTGTACAGAAAGTTGGGT(T)19-3') and the first strand synthesized using Superscript II (Invitrogen). The second strand was synthesized and an attB1 adaptor was ligated to the 5' end (5'-TCGTCGGGACAACTTTGTACAAAGTTGTCCTCC -3' and 5'-CCAACTTTTGTACAAAGTTGTCCTCC -3'). cDNA was size selected using 1% agarose gel electrophoresis (L -0.5-1.2k, M -1.2-2.5k, H ->2.5k) and then inserted into the vector using site specific recombination (flanking attB sites on cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 738;  
Best Local Similarity 87.5%; Pred. No. 9.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
Db 166 AAAAACTGCAATATCAGCCTGAGA 143  
|||||

RESULT 39  
DT108879/c  
LOCUS  
DEFINITION  
JGI\_ANN11096.fwd ANNN Pimephales promelas Whole (L) Pimephales  
promelas cDNA clone ANNN1096 5', mRNA sequence.  
DT108879 743 bp mRNA linear EST 12-AUG-2005  
Pimephales promelas  
Pimephales promelas  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Pimephales.  
1 (bases 1 to 743)  
Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,  
Brokstein, P. and Lindquist, E.A.  
DOE Joint Genome Institute Pimephales promelas EST project  
Unpublished (2005)  
Contact: Lindquist, E.A., Richardson, P.  
Other ESTs: JGI\_ANN11096.rev  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
cDNA Library Preparation: DOE Joint Genome Institute:  
http://www.jgi.doe.gov  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Small Insert: Based upon one or more sequencing reads of this clone  
where vector sequence was present at both ends, this clone has been  
determined to contain a cDNA insert on the order of 600-1000 bases.  
Plate: ANNN 0113 row: p column: 14  
High quality sequence stop: 730.  
Location/Qualifiers  
1..743  
/organism="Pimephales promelas"

FEATURES  
source

/mol\_type="mRNA"  
/db\_xref="taxon:90988"  
/clone="ANN11096"  
/tissue\_type="Whole"  
/dev\_stage="Adult"  
/clone\_lib="ANN Pimephales promelas Whole (L)"  
/note="Vector: pDONR222; The library was made from dT primed cDNA and cloned into Invitrogen vector pDONR222. Poly A RNA were primed with a Biotin-attB2-Oligo(dT) primer (5'-GGCGCGCACAACTTTGTACAGAAAGTTGGGT(T)19-3') and the first strand synthesized using Superscript II (Invitrogen). The second strand was synthesized and an attB1 adaptor was ligated to the 5' end (5'-TCGTCGGGACAACTTTGTACAAAGTTGTCCTCC -3' and 5'-CCAACTTTTGTACAAAGTTGTCCTCC -3'). cDNA was size selected using 1% agarose gel electrophoresis (L -0.5-1.2k, M -1.2-2.5k, H ->2.5k) and then inserted into the vector using site specific recombination (flanking attB sites on cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 743;  
Best Local Similarity 87.5%; Pred. No. 9.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
Db 141 AAAAACTGCAATATCAGCCTGAGA 118  
|||||

RESULT 40  
DT093604/c  
LOCUS  
DEFINITION  
JGI\_ANN725.fwd ANNN Pimephales promelas Whole (L) Pimephales  
promelas cDNA clone ANNN725 5', mRNA sequence.  
DT093604 755 bp mRNA linear EST 12-AUG-2005  
Pimephales promelas  
Pimephales promelas  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Pimephales.  
1 (bases 1 to 755)  
Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,  
Brokstein, P. and Lindquist, E.A.  
DOE Joint Genome Institute Pimephales promelas EST project  
Unpublished (2005)  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
cDNA Library Preparation: DOE Joint Genome Institute:  
http://www.jgi.doe.gov  
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'  
indicates a forward sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Small Insert: Based upon one or more sequencing reads of this clone  
where vector sequence was present at both ends, this clone has been  
determined to contain a cDNA insert on the order of 600-1000 bases.  
Plate: ANNN 0005 row: j column: 14  
High quality sequence stop: 720.  
Location/Qualifiers  
1..755  
/organism="Pimephales promelas"  
/mol\_type="mRNA"  
/db\_xref="taxon:90988"  
/clone="ANN725"

FEATURES  
source

```

/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 755;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||| ||| ||| ||| |||
Db 165 AAAAACTGCAATATCAGCCTGAGA 142

RESULT 41
DT107228/c
LOCUS
DEFINITION
Pimephales cDNA clone ANNN8473 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pimephales promelas
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Pimephales.
1 (bases 1 to 755)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Pimephales promelas EST project
Unpublished (2005)
Other ESTs: JGI ANNN8473.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: ANNN 0089 row: a column: 7
High quality sequence stop: 752.
Location/Qualifiers
1..755
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN8473"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 755;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||| ||| ||| ||| |||
Db 165 AAAAACTGCAATATCAGCCTGAGA 142

RESULT 41
DT107228/c
LOCUS
DEFINITION
Pimephales cDNA clone ANNN8473 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pimephales promelas
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Pimephales.
1 (bases 1 to 755)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Pimephales promelas EST project
Unpublished (2005)
Other ESTs: JGI ANNN8473.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: ANNN 0089 row: a column: 7
High quality sequence stop: 752.
Location/Qualifiers
1..755
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN8473"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 755;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||| ||| ||| ||| |||
Db 167 AAAAACTGCAATATCAGCCTGAGA 144

RESULT 42
DT095581/c
LOCUS
DEFINITION
Pimephales cDNA clone ANNN1867 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pimephales promelas
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Pimephales.
1 (bases 1 to 769)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Pimephales promelas EST project
Unpublished (2005)
Other ESTs: JGI ANNN1867.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: ANNN 0017 row: f column: 12
High quality sequence stop: 761.
Location/Qualifiers
1..769
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN1867"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

FEATURES
source
Location/Qualifiers
1..769
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN1867"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

```

```

/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 755;
Best Local Similarity 87.5%; Pred. No. 9.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||| ||| ||| ||| ||| |||
Db 167 AAAAACTGCAATATCAGCCTGAGA 144

RESULT 42
DT095581/c
LOCUS
DEFINITION
Pimephales cDNA clone ANNN1867 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pimephales promelas
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Pimephales.
1 (bases 1 to 769)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Pimephales promelas EST project
Unpublished (2005)
Other ESTs: JGI ANNN1867.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: ANNN 0017 row: f column: 12
High quality sequence stop: 761.
Location/Qualifiers
1..769
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN1867"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

FEATURES
source
Location/Qualifiers
1..769
/organism="Pimephales promelas"
/mol_type="mRNA"
/db_xref="taxon:90988"
/clone="ANNN1867"
/tissue_type="Whole"
/dev_stage="Adult"
/clone_lib="ANNN Pimephales promelas Whole (L)"
/notes="Vector: pDONR222; The library was made from dT
primed cDNA and cloned into Invitrogen vector pDONR222.
Poly A RNA were primed with a Biotin-attB2-Oligo(dT)
primer (5'-GGGGCGGCACAACTTTGTACAAGAAAGTGGGT(T)19-3')
and the first strand synthesized using Superscript II
(invitrogen). The second strand was synthesized and an
atBI adaptor was ligated to the 5' end (5'-
TCGTGGGACAACTTTGTACAAAAGTTGG-3' and 5'-
CCAACCTTTTGTACAAGTTGTCCTCC-3'). cDNA was size selected
using 1% agarose gel electrophoresis (L -0.5-1.2k, M
-1.2-2.5k, H -2.5k) and then inserted into the vector
using site specific recombination (flanking attB sites on
cDNA). The work was done at DOE Joint Genome Institute."

```



-1.2-2.5k, H ->2.5k) and then inserted into the vector using site specific recombination (flanking attB sites on cDNA). The work was done at DOE Joint Genome Institute."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 791;  
Best Local Similarity 87.5%; Pred. No. 9.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
DB 167 AAAAACTGGAATCTCAGGCTGAGA 144  
|||||

## RESULT 45

BX241295/c 791 bp DNA linear GSS 13-MAR-2003  
LOCUS Danio rerio genomic clone DKEY-246B9, genomic survey sequence.  
DEFINITION BX241295  
ACCESSION BX241295  
VERSION BX241295.1 GI:28163629  
KEYWORDS GSS.  
SOURCE Danio rerio (zebrafish)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE Humphray, S.J., Huckle, E. and Durham, J.L.  
AUTHORS Direct Submission  
TITLE Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome  
JOURNAL Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Unpublished

COMMENT This sequence was generated from the T7 end of BAC 246B9. 246B9 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:  
http://www.sanger.ac.uk/Projects/D\_rerio/.

## FEATURES

source  
1..791  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-246B9"  
/tissue types="Testis"  
/note="Vector pindigoBAC-536"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 791;  
Best Local Similarity 87.5%; Pred. No. 9.7e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
DB 343 AAAAACTGGAATCCAGCTCTTGA 320  
|||||

## RESULT 46

BZ452080 824 bp DNA linear GSS 13-DEC-2002  
LOCUS BOND84TF BO 1.6\_2 KB tot Brassica oleracea genomic clone BOND84,  
DEFINITION genomic survey sequence.  
ACCESSION BZ452080  
VERSION BZ452080.1 GI:26724809  
KEYWORDS GSS.  
SOURCE Brassica oleracea

ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 824)

AUTHORS Ayele, M., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,  
Utterback, T.R., Wortman, J.R., White, O.R. and Town, C.D.  
TITLE Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis

JOURNAL PUBMED  
COMMENT

Genome Res. 15 (4), 487-495 (2005)  
15805490  
Other GSSs: BOND84TR  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source

Location/Qualifiers  
1..824  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOND84"  
/clone\_lib="BO\_1.6\_2\_KB\_tot"  
/note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 824;  
Best Local Similarity 87.5%; Pred. No. 9.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
DB 524 AAAAACTGGAATCTCAGGCTGAGA 547  
|||||

## RESULT 47

BUI25751 861 bp mRNA linear EST 25-NOV-2002  
LOCUS 603149763F1 CSEQCHL19 Gallus gallus cDNA clone CHEST154n16 5', mRNA  
DEFINITION sequence.  
ACCESSION BUI25751  
VERSION BUI25751.1 GI:25336704  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 861)  
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
PUBMED 12445392  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
source

Location/Qualifiers  
1..861  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="CHEST154n16"  
/sex="Female"  
/dev stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHL19"  
/note="Organ: liver; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)

[Stratagene] vector to accommodate cDNA produced with the T-trimmed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BsgI and BamHI sites [5'ggcgcgtcagcccgatccgaaaaag] [5'aattcttttttcggatccggggtcagcgc]"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 5; Length 861;  
Best Local Similarity 87.5%; Pred. No. 9.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGAGA 25  
|||||  
DB 829 AAAAAGTGAATCACAGTCTGAAA 852

## RESULT 48

CD126154 508 bp mRNA linear EST 14-SEP-2003  
LOCUS MF1-0002U-V053-E04-U.B MF1-0002 Schistosoma mansoni cDNA clone  
DEFINITION MF1-0002U-V053-E04.B, mRNA sequence.  
ACCESSION CD126154  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

## REFERENCE

1 (bases 1 to 508)  
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr., Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F., Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L., Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A., Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A., Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M., Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.  
Transcriptome analysis of the acelomate human parasite Schistosoma mansoni  
Nat. Genet. 35 (2), 148-157 (2003)  
Contact: Dr. Sergio Verjovski-Almeida  
Departamento de Bioquímica  
Instituto de Química - Universidade de São Paulo  
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP, Brazil  
Tel: +55-11-3091-2173  
Fax: +55-11-3091-2186  
Email: verj@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>  
Plate: MF1-0002U-V053 row: 4 column: E.

## FEATURES

1. 508  
Location/Qualifiers  
/organism="Schistosoma mansoni"  
/mol\_type="mRNA"  
/db\_xref="taxon:6183"  
/clone="MF1-0002U-V053-E04.B"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="Mesocricetus auratus"  
/clone\_lib="MF1-0002"  
/note="Vector: SureClone"

## ORIGIN

Query Match 76.0%; Score 19; DB 6; Length 508;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AAAAAGTGAATCTCAGGCTGAGA 20  
|||||  
DB 481 AAAAAGTGAATCTCAGGCTGAGA 499

## RESULT 49

BU254969 634 bp mRNA linear EST 26-NOV-2002  
LOCUS 603414046F1 CSEQCHN38 Gallus gallus cDNA clone CHEST335111 5', mRNA  
DEFINITION  
ACCESSION BU254969  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

EST: BU254969.1 GI:25511230  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 634)

## REFERENCE

BOARDMAN, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

## COMMENT

Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

1. 634  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST335111"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN38"  
/note="Organ: limbs; Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

## ORIGIN

Query Match 76.0%; Score 19; DB 5; Length 634;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AAAAAGTGAATCTCAGGCTG 22  
|||||  
DB 490 AAAAAGTGAATCTCAGGCTG 508

## RESULT 50

BU338343 659 bp mRNA linear EST 28-NOV-2002  
LOCUS 603515585F1 CSEQCHN66 Gallus gallus cDNA clone CHEST454j22 5', mRNA  
DEFINITION

sequence.  
ACCESSION BU338343  
VERSION BU338343.1 GI:25846344  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 659)  
REFERENCE Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
12445392  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers  
FEATURES  
Source 1..659  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST454j22"  
/dev\_stage="16 day embryo"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN66"  
/notes="Organ: Brain; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

ORIGIN  
Query Match 76.0%; Score 19; DB 5; Length 659;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 AACTGGAAATCTCAGGCTG 22  
|||||  
Db 281 AACTGGAAATCTCAGGCTG 299

Search completed: February 3, 2006, 22:02:39  
Job time : 2959.67 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-30

Perfect score: 25

Sequence: 1 aaaaactggaatcaggtgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	77.6	39552	3	US-09-949-016-11799
2	19.2	76.8	11003	3	US-09-949-016-13166
3	18.8	75.2	55195	3	US-09-949-016-15854
4	18.6	74.4	192506	3	US-09-949-016-15830
5	18.2	72.8	95318	3	US-09-949-016-11784
6	18.2	72.8	95318	3	US-09-949-016-13908
7	17.8	71.2	601	3	US-09-949-016-45703
8	17.8	71.2	150394	3	US-09-949-016-13042
9	17.6	70.4	601	3	US-09-949-016-48777
10	17.6	70.4	31318	3	US-09-949-016-12495
11	17.6	70.4	31319	3	US-09-949-016-15963
12	17.6	70.4	51928	3	US-09-949-016-13184
13	17.2	68.8	9300	3	US-09-949-016-12482
14	17.2	68.8	9301	3	US-09-949-016-17288
15	17.2	68.8	26619	3	US-09-949-016-15030
16	17.2	68.8	41612	3	US-09-949-016-12769
17	17.2	68.8	41639	3	US-09-949-016-12773
18	17.2	68.8	77388	3	US-09-949-016-13496
19	17.2	68.8	112705	3	US-09-949-016-15630
20	17.2	68.8	115388	3	US-09-949-016-14981
21	17.2	68.8	127771	3	US-09-949-016-14982
22	17.2	68.8	264206	3	US-09-949-016-12731
23	17.2	68.8	264304	3	US-09-949-016-13249
24	17	68.0	234	3	US-09-248-796A-7364

25	17	68.0	304	3	US-09-513-999C-24054	Sequence 24054, A
26	17	68.0	601	3	US-09-949-016-145029	Sequence 145029, A
27	17	68.0	601	3	US-09-949-016-145030	Sequence 145030, A
28	17	68.0	601	3	US-09-949-016-170085	Sequence 170085, A
29	17	68.0	631	3	US-09-513-999C-14974	Sequence 14974, A
30	17	68.0	1111	3	US-09-167-206-9	Sequence 9, Appli
31	17	68.0	1138	3	US-10-164-595-33	Sequence 3, Appli
32	17	68.0	1424	3	US-08-956-171B-566	Sequence 566, App
33	17	68.0	1424	3	US-08-781-986A-566	Sequence 566, App
34	17	68.0	1717	3	US-08-956-171B-481	Sequence 481, App
35	17	68.0	1717	3	US-08-781-986A-481	Sequence 481, App
36	17	68.0	1858	3	US-09-347-819-1	Sequence 1, Appli
37	17	68.0	2880	3	US-10-104-047-859	Sequence 859, App
38	17	68.0	7736	3	US-09-949-016-17579	Sequence 17579, A
39	17	68.0	20347	3	US-09-949-016-16752	Sequence 16752, A
40	17	68.0	22965	3	US-09-949-016-15862	Sequence 15862, A
41	17	68.0	23533	3	US-09-949-016-15377	Sequence 15377, A
42	17	68.0	42989	3	US-09-949-016-12291	Sequence 12291, A
43	17	68.0	42992	3	US-09-949-016-15428	Sequence 15428, A
44	17	68.0	49677	3	US-09-949-002-817	Sequence 817, App
45	17	68.0	49818	3	US-09-949-016-12458	Sequence 12458, A
46	17	68.0	49829	3	US-09-949-016-14011	Sequence 14011, A
47	17	68.0	85675	3	US-09-949-016-12333	Sequence 12333, A
48	17	68.0	85675	3	US-09-949-016-15956	Sequence 15956, A
49	17	68.0	96327	3	US-09-949-016-16541	Sequence 16541, A
50	17	68.0	105189	3	US-09-949-016-13029	Sequence 13029, A
51	17	68.0	129327	3	US-09-949-016-12257	Sequence 12257, A
52	17	68.0	129327	3	US-09-949-016-15368	Sequence 15368, A
53	17	68.0	177251	3	US-09-949-016-15941	Sequence 15941, A
54	17	68.0	194790	3	US-09-949-016-15393	Sequence 15393, A
55	17	68.0	260247	3	US-09-949-016-13358	Sequence 13358, A
56	17	68.0	1230025	3	US-09-198-452A-1	Sequence 1, Appli
57	17	68.0	1230230	3	US-09-438-185A-1	Sequence 1, Appli
58	16.8	67.2	133	3	US-09-513-999C-27290	Sequence 27290, A
59	16.8	67.2	601	3	US-09-949-016-59983	Sequence 59983, A
60	16.8	67.2	601	3	US-09-949-016-117283	Sequence 117283, A
61	16.8	67.2	601	3	US-09-949-016-117352	Sequence 117352, A
62	16.8	67.2	31214	3	US-09-949-016-17097	Sequence 17097, A
63	16.8	67.2	31214	3	US-09-949-016-17098	Sequence 17098, A
64	16.8	67.2	32958	3	US-09-949-016-15677	Sequence 15677, A
65	16.8	67.2	36242	3	US-09-949-016-12986	Sequence 12986, A
66	16.8	67.2	36242	3	US-09-949-016-12997	Sequence 12997, A
67	16.8	67.2	36242	3	US-09-949-016-12998	Sequence 12998, A
68	16.8	67.2	36242	3	US-09-949-016-12999	Sequence 12999, A
69	16.8	67.2	36242	3	US-09-949-016-13000	Sequence 13000, A
70	16.8	67.2	46526	3	US-09-949-016-13390	Sequence 13390, A
71	16.8	67.2	60595	3	US-09-949-002-650	Sequence 650, App
72	16.8	67.2	60595	3	US-09-949-002-706	Sequence 706, App
73	16.8	67.2	83450	3	US-09-811-469-3	Sequence 3, Appli
74	16.8	67.2	83450	3	US-10-370-659-3	Sequence 3, Appli
75	16.8	67.2	168394	3	US-09-949-016-13002	Sequence 13002, A
76	16.6	66.4	217	3	US-09-513-999C-3132	Sequence 3132, Ap
77	16.6	66.4	472	3	US-09-270-767-6329	Sequence 6329, Ap
78	16.6	66.4	472	3	US-09-270-767-6329	Sequence 21611, A
79	16.6	66.4	601	3	US-09-949-016-53654	Sequence 53654, A
80	16.6	66.4	601	3	US-09-949-016-80518	Sequence 80518, A
81	16.6	66.4	651	3	US-09-270-767-7426	Sequence 7426, Ap
82	16.6	66.4	651	3	US-09-270-767-22708	Sequence 22708, A
83	16.6	66.4	858	3	US-09-621-976-3699	Sequence 3699, Ap
84	16.6	66.4	885	3	US-09-270-767-8281	Sequence 8281, Ap
85	16.6	66.4	885	3	US-09-270-767-23563	Sequence 23563, A
86	16.6	66.4	1308	3	US-09-248-796A-2964	Sequence 2964, App
87	16.6	66.4	1990	3	US-09-149-476-281	Sequence 281, App
88	16.6	66.4	2041	3	US-09-149-476-131	Sequence 131, App
89	16.6	66.4	2334	2	US-08-062-632-4	Sequence 4, Appli
90	16.6	66.4	26173	3	US-09-949-016-14422	Sequence 14422, A
91	16.6	66.4	26173	3	US-09-453-702B-69	Sequence 69, Appli
92	16.6	66.4	34011	3	US-10-114-170-69	Sequence 69, Appli
93	16.6	66.4	34011	3	US-09-949-016-12485	Sequence 12485, A
94	16.6	66.4	36618	3	US-09-949-016-12961	Sequence 12961, A
95	16.6	66.4	48908	3	US-09-453-702B-137	Sequence 137, App
96	16.6	66.4	48908	3	US-10-114-170-137	Sequence 137, App
97	16.6	66.4	60137	3	US-09-949-016-14735	Sequence 14735, A





```
Best Local Similarity 90.5%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGA 23
    ||||| ||||| ||||| |||||
Db 46 AAAAAGTGAATCTCAGGCTGA 66

RESULT 8
US-09-949-016-13042/c
; Sequence 13042, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13042
; LENGTH: 150394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(150394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13042

Query Match 71.2%; Score 17.8; DB 3; Length 150394;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGA 23
    ||||| ||||| ||||| |||||
Db 124943 AAAAAGTGAATCTCAGGCTGA 124923

RESULT 9
US-09-949-016-48777/c
; Sequence 48777, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48777
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48777

Query Match 70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGA 25
    ||||| ||||| ||||| |||||
Db 249 AAAAAGGGGACACTCAGGCTGA 226

RESULT 10
US-09-949-016-12495
; Sequence 12495, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12495
; LENGTH: 31318
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12495

Query Match 70.4%; Score 17.6; DB 3; Length 31318;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGTGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| |||||
Db 11006 AAAAAGTGAATCTCAGGCTGAG 11029

RESULT 11
US-09-949-016-15963
; Sequence 15963, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15963
; LENGTH: 31319
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15963

Query Match 70.4%; Score 17.6; DB 3; Length 31319;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGTGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| |||||
Db 11006 AAAAAGTGAATCTCAGGCTGAG 11029
```



```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12769
; LENGTH: 41612
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(41612)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12769

```

```
Query Match          68.8%; Score 17.2; DB 3; Length 41612;
Best Local Similarity 86.4%; Pred. No. 3.1e+02;
Matches 19: Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 4 AAACCTGGAAATCTCAGGCTGAGA 25  
1894 AACCTGGAAATCTCAGTCTAAGA 1915  
Db

```

RESULT 17
US-09-949-016-12773
; Sequence 12773, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12773
; LENGTH: 41639
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(41639)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12773

```

Query Match 68.8%; Score 17.2; DB 3; Length 41639;  
Best Local Similarity 86.4%; Pred. No. 3.1e+02;  
Matches 19: Conservative 0; Mismatches 3: Indels 0;

QY 4 AAACCTGGAATCTCAGGCTGAGA 25  
19921 AACCTGGAATCTCAGTCTAAGA 19942

RESULT 18

```

US-09-949-016-13496
; Sequence 13496, Application US/09949016
; Patent NO. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13496
; LENGTH: 77388
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13496

```

Query Match	68.8%	Score 17.2;	DB 3;	Length 77388;
Best Local Similarity	86.4%	Pred. No. 3.4e+02;		
Matches 19: Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY  
1 AAAAACTGGAATCTCAGGCTG 22  
|||  
74064 AAAGAACTGGAAATTTCAGGATG 74085

Dβ

```

RESULT 19
US-09-949-016-15630/c
; Sequence 15630, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15630
; LENGTH: 112705
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(112705)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15630

```

```
Query Match      68.8%; Score 17.2; DB 3; Length 112705;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19: Conservative 0; Mismatches 3; Indels 0; Gaps 0
```

Qy 1 AAAAACTGGAATCTCAGGCTG 22  
||| ||| ||| ||| ||| |||  
Db 60506 AAATAACTGGAATGTGAGGCTG 60485

RESULT 20  
US-09-949-016-14981

```

; Sequence 14981, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14981
; LENGTH: 115388
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14981

Query Match      68.8%; Score 17.2; DB 3; Length 115388;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTG 22
Db 74064 AAAGAACTGGAATTCAGGATG 74085

RESULT 21
US-09-949-016-14982
; Sequence 14982, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14982
; LENGTH: 127771
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14982

Query Match      68.8%; Score 17.2; DB 3; Length 127771;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTG 22
Db 74064 AAAGAACTGGAATTCAGGATG 74085

RESULT 22
US-09-949-016-12731/c
; Sequence 12731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12731
; LENGTH: 264206
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12731

Query Match      68.8%; Score 17.2; DB 3; Length 264206;
Best Local Similarity 86.4%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGAGA 25
Db 245257 AAACCTGGCATCTCTGGCTAAGA 245236

RESULT 23
US-09-949-016-13249/c
; Sequence 13249, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13249
; LENGTH: 264304
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13249

Query Match      68.8%; Score 17.2; DB 3; Length 264304;
Best Local Similarity 86.4%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGAGA 25
Db 245257 AAACCTGGCATCTCTGGCTAAGA 245236

RESULT 24
US-09-248-796A-7364
; Sequence 7364, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725

```





```
RESULT 28
US-09-949-016-170085/c
; Sequence 170085, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170085
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170085

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 182 AAAATGTGAAACTGAGGCTTAGA 158

RESULT 29
US-09-513-999C-14974
; Sequence 14974, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.Pm
; SEQ ID NO 14974
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 137
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 138
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 182
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 201
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: 356
; OTHER INFORMATION: m=a or c
US-09-513-999C-14974

Query Match      68.0%; Score 17; DB 3; Length 631;
Best Local Similarity 80.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 516 AAGTCACTGGAGGCTCAGGCTGAGA 540

RESULT 30
US-09-167-206-9
; Sequence 9, Application US/09167206A
; Patent No. 6476193
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Schulz, Vincent P.
; APPLICANT: Yang, Meijia
; TITLE OF INVENTION: Nlki PROTEIN AND Nlki PROTEIN COMPLEXES
; FILE REFERENCE: 15966-521 Nlki protein complexes
; CURRENT APPLICATION NUMBER: US/09/167,206A
; CURRENT FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 111
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (57)..(908)
US-09-167-206-9

Query Match      68.0%; Score 17; DB 3; Length 111;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 669 AAGTCACTGGAGGCTCAGGCTGAGA 693

RESULT 31
US-10-164-595-33
; Sequence 33, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1138
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(965)
; OTHER INFORMATION:
US-10-164-595-33

Query Match      68.0%; Score 17; DB 3; Length 1138;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25
```

Db	735 AAGTCACTGGAGGCTCAGGCTGAGA 759
RESULT 32	
US-08-956-171E-566/c	
Sequence 566, Application US/08956171E	
Patent No. 6593114	
GENERAL INFORMATION:	
APPLICANT: Charles Kunsch	
Gil H. Choi	
Patrick S. Dillon	
Craig A. Rosen	
Steven C. Barash	
Michael R. Fannon	
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences	
NUMBER OF SEQUENCES: 5256	
CORRESPONDENCE ADDRESS:	
ADDRESSEE: Human Genome Sciences, Inc.	
STREET: 9410 Key West Avenue	
CITY: Rockville	
STATE: Maryland	
COUNTRY: USA	
ZIP: 20850	
COMPUTER READABLE FORM:	
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage	
COMPUTER: HP Vectra 486/33	
OPERATING SYSTEM: MSDOS version 6.2	
SOFTWARE: ASCII Text	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/956,171E	
FILING DATE:	
CLASSIFICATION: 435	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER:	
FILING DATE:	
ATTORNEY/AGENT INFORMATION:	
NAME: Benson, Bob	
REGISTRATION NUMBER: 30,446	
REFERENCE/DOCKET NUMBER: PB248PP	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (301) 309-8504	
TELEFAX: (301) 309-8512	
INFORMATION FOR SEQ ID NO: 566:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1424 base pairs	
TYPE: nucleic acid	
STRANDEDNESS: double	
TOPOLOGY: linear	
US-08-781-986A-566	
Query Match	68.0%; Score 17; DB 3; Length 1424;
Best Local Similarity	80.0%; Pred. No. 2.2e+02;
Matches	20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	1 AAAAACTGGAACTCAGGCTGAGA 25
Db	385 AACAAATCGGAATCTAATGCTGAGA 361
RESULT 34	
US-08-956-171E-481	
Sequence 481, Application US/08956171E	
Patent No. 6593114	
GENERAL INFORMATION:	
APPLICANT: Charles Kunsch	
Gil H. Choi	
Patrick S. Dillon	
Craig A. Rosen	
Steven C. Barash	
Michael R. Fannon	
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences	
NUMBER OF SEQUENCES: 5256	
CORRESPONDENCE ADDRESS:	
ADDRESSEE: Human Genome Sciences, Inc.	
STREET: 9410 Key West Avenue	
CITY: Rockville	
STATE: Maryland	
COUNTRY: USA	
ZIP: 20850	
COMPUTER READABLE FORM:	
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage	
COMPUTER: HP Vectra 486/33	
OPERATING SYSTEM: MSDOS version 6.2	
SOFTWARE: ASCII Text	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/956,171E	
FILING DATE: 20-Oct-1997	
CLASSIFICATION: <Unknown>	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER: 60/009,861	
FILING DATE: January 5, 1996	
APPLICATION NUMBER: 08/781,986	
FILING DATE: January 3, 1997	
ATTORNEY/AGENT INFORMATION:	
NAME: Mark J. Hyman	
REGISTRATION NUMBER: 46,789	
REFERENCE/DOCKET NUMBER: PB248P1	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (240) 314-1224	
TELEFAX: (301) 309-8439	
INFORMATION FOR SEQ ID NO: 566:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1424 base pairs	
TYPE: nucleic acid	
STRANDEDNESS: double	
TOPOLOGY: linear	
US-08-956-171E-566	
Query Match	68.0%; Score 17; DB 3; Length 1424;
Best Local Similarity	80.0%; Pred. No. 2.2e+02;
Matches	20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	1 AAAAACTGGAACTCAGGCTGAGA 25
Db	385 AACAAATCGGAATCTAATGCTGAGA 361
RESULT 33	
US-08-781-986A-566/c	
Sequence 566, Application US/08781986A	
Patent No. 6737248	
GENERAL INFORMATION:	
APPLICANT: Charles Kunsch	
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences	
NUMBER OF SEQUENCES: 5255	
CORRESPONDENCE ADDRESS:	
ADDRESSEE: Human Genome Sciences, Inc.	

APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 481:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1717 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 481:  
US-08-956-171E-481

Query Match . 68.0%; Score 17; DB 3; Length 1717;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 817 AAAAACTGGAATCTCAGGCTTAGA 841

RESULT 35  
US-08-781-986A-481  
; Sequence 481, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 481:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1717 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-481

Query Match 68.0%; Score 17; DB 3; Length 1717;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 817 AAAAACTGGAATCTCAGGCTTAGA 841

RESULT 36  
US-09-347-819-1  
; Sequence 1, Application US/09347819  
; Patent No. 6184036  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Rebecca E.  
; APPLICANT: Rafalski, J. Antoni  
; APPLICANT: Falco, S. Carl  
; TITLE OF INVENTION: Ornithine Biosynthesis Enzymes  
; FILE REFERENCE: BB-1174-C  
; CURRENT APPLICATION NUMBER: US/09/347,819  
; CURRENT FILING DATE: 1999-07-02  
; EARLIER APPLICATION NUMBER: 60/093,209  
; EARLIER FILING DATE: July 17, 1998  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 1  
; LENGTH: 1858  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (10)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1138)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1175)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1201)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1342)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1344)  
US-09-347-819-1

Query Match 68.0%; Score 17; DB 3; Length 1858;  
Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 1557 AAAAACTGGAATCTCAGGCTGAGA 1581

RESULT 37  
US-10-104-047-859  
; Sequence 859, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 859  
; LENGTH: 2880  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-859



```

Db      18450 AAGAACTAGAAATCGAGGCTCAGA 18474

RESULT 44
US-09-949-002-817
; Sequence 817, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 817
; LENGTH: 49677
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-817

Query Match      68.0%; Score 17; DB 3; Length 49677;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      1 AAAAACTGGAATCTCAGGCTGAGA 25
      ||||| ||||| ||||| ||||| |||||
Db      37426 AAAAAATGGAACTGAGGCTAGGA 37450

RESULT 45
US-09-949-016-12458/c
; Sequence 12458, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12458
; LENGTH: 49818
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12458

Query Match      68.0%; Score 17; DB 3; Length 49818;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      1 AAAAACTGGAATCTCAGGCTGAGA 25
      ||||| ||||| ||||| ||||| |||||
Db      36549 AAAAACTAAATCTACGCTGAGA 36525

RESULT 46
US-09-949-016-14011/c
; Sequence 14011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```

```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14011
; LENGTH: 49829
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14011

Query Match      68.0%; Score 17; DB 3; Length 49829;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAACTGGAATCTCAGGCTGAGA 25
DB      36549 AAAAACTGGAATCTCAGGCTGAGA 36525

RESULT 47
US-09-949-016-12333
; Sequence 12333, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12333
; LENGTH: 85675
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12333

Query Match      68.0%; Score 17; DB 3; Length 85675;
Best Local Similarity 80.0%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAACTGGAATCTCAGGCTGAGA 25
DB      55721 AAAAACTGGAATCTCAGGCTGAGA 55745

RESULT 48
US-09-949-016-15956
; Sequence 15956, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15956
; LENGTH: 85675
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15956

Query Match      68.0%; Score 17; DB 3; Length 85675;
Best Local Similarity 80.0%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAACTGGAATCTCAGGCTGAGA 25
DB      55721 AAAAACTGGAATCTCAGGCTGAGA 55745

RESULT 49
US-09-949-016-16541/C
; Sequence 16541, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16541
; LENGTH: 96327
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...(96327)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16541

Query Match      68.0%; Score 17; DB 3; Length 96327;
Best Local Similarity 80.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAACTGGAATCTCAGGCTGAGA 25
DB      15778 AAAAACTGGAATCTCAGGCTGAGA 15754

RESULT 50
US-09-949-016-13029
; Sequence 13029, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```



; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13029  
; LENGTH: 105189  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(105189)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13029

Query Match 68.0%; Score 17; DB 3; Length 105189;  
Best Local Similarity 80.0%; Pred. No. 4.3e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25  
|||||  
Db 17473 AAAAACTGGAATCTCAGGCTGAGA 17497  
|||||

Search completed: February 3, 2006, 16:32:49  
Job time : 88.6667 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-30

Perfect score: 25

Sequence: 1 aaaaactggaatctcaggctgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	8	US-10-719-900-30
2	23.4	93.6	25	8	US-10-719-900-29
3	20.2	80.8	583	4	US-09-925-065A-401109
4	20	80.0	300	9	US-10-779-543-1061
5	20	80.0	758	9	US-10-779-543-3989
6	18.8	75.2	537	4	US-09-925-065A-265072
7	18.8	75.2	63693	7	US-10-741-601-5650
8	18.6	74.4	589	4	US-09-925-065A-517136
9	18.6	74.4	2066	6	US-10-094-749-471
10	18.4	73.6	633	4	US-09-925-065A-417056
11	18.2	72.8	623	4	US-09-925-065A-637551
12	18.2	72.8	633	4	US-09-925-065A-523476
13	18.2	72.8	633	4	US-09-925-065A-523477
14	18.2	72.8	823	5	US-10-027-632-33065
15	18.2	72.8	823	6	US-10-027-632-33066
16	18.2	72.8	823	6	US-10-027-632-33065
17	18.2	72.8	823	6	US-10-027-632-33066
18	17.8	71.2	525	4	US-09-925-065A-362629
19	17.8	71.2	525	4	US-09-925-065A-362630
20	17.8	71.2	525	4	US-09-925-065A-362631
21	17.8	71.2	575	5	US-10-027-632-202264
22	17.8	71.2	575	6	US-10-027-632-202264
23	17.8	71.2	876	7	US-10-425-114-13369

7	884	71.2	17.8	US-10-424-599-113563	Sequence 113563,
1911	201	71.2	17.8	US-10-369-493-24074	Sequence 24074, A
6	201	70.4	17.6	US-10-741-600-53948	Sequence 53948, A
214	214	70.4	17.6	US-09-933-797-614	Sequence 614, App
383	383	70.4	17.6	US-10-242-535A-41892	Sequence 41892, A
383	383	70.4	17.6	US-10-085-789A-41892	Sequence 41892, A
514	514	70.4	17.6	US-09-925-065A-311398	Sequence 311398,
516	516	70.4	17.6	US-09-925-065A-150366	Sequence 150366,
522	522	70.4	17.6	US-10-029-386-2155	Sequence 2155, Ap
538	538	70.4	17.6	US-09-925-065A-501583	Sequence 501583,
549	549	70.4	17.6	US-09-925-065A-435459	Sequence 435459,
549	549	70.4	17.6	US-09-925-065A-435470	Sequence 435470,
549	549	70.4	17.6	US-09-925-065A-435471	Sequence 435471,
549	549	70.4	17.6	US-09-925-065A-435473	Sequence 435473,
600	600	70.4	17.6	US-11-060-756-455	Sequence 455, App
600	600	70.4	17.6	US-11-060-756-456	Sequence 456, App
600	600	70.4	17.6	US-11-060-756-4727	Sequence 4727, Ap
604	604	70.4	17.6	US-11-060-756-4728	Sequence 4728, Ap
604	604	70.4	17.6	US-09-925-065A-288955	Sequence 288955,
606	606	70.4	17.6	US-09-925-065A-265864	Sequence 265864,
631	631	70.4	17.6	US-09-925-065A-563312	Sequence 563312,
656	656	70.4	17.6	US-10-027-632-191799	Sequence 191799,
656	656	70.4	17.6	US-10-027-632-191799	Sequence 191799,
1469	1469	70.4	17.6	US-09-925-065A-721185	Sequence 721185,
1551	1551	70.4	17.6	US-10-424-599-123483	Sequence 123483,
5271	5271	70.4	17.6	US-10-044-090-354	Sequence 354, App
5271	5271	70.4	17.6	US-10-084-817-340	Sequence 340, App
24526	24526	70.4	17.6	US-09-749-589-3	Sequence 3, Appli
24526	24526	70.4	17.6	US-10-684-532-3	Sequence 3, Appli
26320	26320	70.4	17.6	US-10-166-221-3	Sequence 3, Appli
37996	37996	70.4	17.6	US-11-097-143-9055	Sequence 9055, Ap
51917	51917	70.4	17.6	US-10-741-600-17758	Sequence 17758, A
96595	96595	70.4	17.6	US-09-997-722-43	Sequence 43, Appl
154902	154902	70.4	17.6	US-10-450-826-105	Sequence 105, App
299598	299598	70.4	17.6	US-10-322-696-16	Sequence 16, Appl
350764	350764	70.4	17.6	US-10-087-192-1864	Sequence 1864, Ap
653122	653122	70.4	17.6	US-10-087-192-226	Sequence 226, App
580	580	69.6	17.4	US-09-925-065A-474511	Sequence 474511,
591	591	69.6	17.4	US-09-925-065A-783432	Sequence 783432,
739	739	69.6	17.4	US-10-027-632-169456	Sequence 169456,
739	739	69.6	17.4	US-10-027-632-169456	Sequence 169456,
1059	1059	69.6	17.4	US-10-425-114-219	Sequence 219, App
2206	2206	69.6	17.4	US-10-425-114-2686	Sequence 2686, Ap
2800	2800	69.6	17.4	US-10-425-114-34756	Sequence 34756, A
2899	2899	69.6	17.4	US-10-425-115-143582	Sequence 143582,
35456	35456	69.6	17.4	US-10-087-192-103	Sequence 103, App
25	25	68.8	17.2	US-11-036-317-159236	Sequence 159236,
25	25	68.8	17.2	US-11-036-317-217422	Sequence 217422,
395	395	68.8	17.2	US-09-925-065A-604141	Sequence 604141,
428	428	68.8	17.2	US-09-925-065A-167140	Sequence 167140,
460	460	68.8	17.2	US-09-732-627A-3090	Sequence 3090, Ap
516	516	68.8	17.2	US-09-925-065A-150367	Sequence 150367,
518	518	68.8	17.2	US-09-925-065A-113663	Sequence 113663,
518	518	68.8	17.2	US-09-925-065A-113664	Sequence 113664,
518	518	68.8	17.2	US-09-925-065A-113665	Sequence 113665,
519	519	68.8	17.2	US-09-925-065A-522586	Sequence 522586,
519	519	68.8	17.2	US-09-925-065A-522588	Sequence 522588,
521	521	68.8	17.2	US-09-925-065A-647331	Sequence 647331,
533	533	68.8	17.2	US-10-027-632-180346	Sequence 180346,
533	533	68.8	17.2	US-10-027-632-299920	Sequence 299920,
533	533	68.8	17.2	US-10-027-632-180346	Sequence 180346,
533	533	68.8	17.2	US-10-027-632-299920	Sequence 299920,
536	536	68.8	17.2	US-10-027-632-76404	Sequence 76404, A
536	536	68.8	17.2	US-10-027-632-76404	Sequence 76404, A
546	546	68.8	17.2	US-10-027-632-228656	Sequence 228656,
546	546	68.8	17.2	US-10-027-632-228656	Sequence 228656,
547	547	68.8	17.2	US-09-925-065A-358845	Sequence 358845,
559	559	68.8	17.2	US-09-925-065A-14018	Sequence 14018, A
560	560	68.8	17.2	US-09-925-065A-312445	Sequence 312445,
569	569	68.8	17.2	US-09-925-065A-221346	Sequence 221346,
569	569	68.8	17.2	US-10-027-632-286926	Sequence 286926,
569	569	68.8	17.2	US-10-027-632-286926	Sequence 286926,
572	572	68.8	17.2	US-09-925-065A-221345	Sequence 221345,

c 97	17.2	68.8	573	5	US-10-027-632-84863	Sequence 84863, A
	17.2	68.8	573	5	US-10-027-632-84864	Sequence 84864, A
c 98	17.2	68.8	573	5	US-10-027-632-286054	Sequence 286054, A
c 99	17.2	68.8	573	5	US-10-027-632-286055	Sequence 286055, A
c 100	17.2	68.8	573	5	US-10-027-632-84863	Sequence 84863, A
c 101	17.2	68.8	573	6	US-10-027-632-84864	Sequence 84864, A
c 102	17.2	68.8	573	6	US-10-027-632-286054	Sequence 286054, A
c 103	17.2	68.8	573	6	US-10-027-632-286055	Sequence 286055, A
c 104	17.2	68.8	573	6	US-10-027-632-286055	Sequence 286055, A
c 105	17.2	68.8	589	4	US-09-925-065A-756159	Sequence 756159, A
c 106	17.2	68.8	598	4	US-09-925-065A-883540	Sequence 883540, A
c 107	17.2	68.8	604	4	US-09-925-065A-426028	Sequence 426028, A
c 108	17.2	68.8	664	5	US-10-027-632-242180	Sequence 242180, A
c 109	17.2	68.8	664	6	US-10-027-632-242180	Sequence 242180, A
c 110	17.2	68.8	672	4	US-09-925-065A-670810	Sequence 670810, A
c 111	17.2	68.8	672	4	US-09-925-065A-670811	Sequence 670811, A
c 112	17.2	68.8	693	8	US-10-425-115-15337	Sequence 15337, A
c 113	17.2	68.8	877	9	US-10-779-543-4813	Sequence 4813, Ap
c 114	17.2	68.8	944	8	US-10-767-795-5251	Sequence 5251, Ap
c 115	17.2	68.8	1184	4	US-09-925-065A-699263	Sequence 699263, A
c 116	17.2	68.8	1448	8	US-10-767-795-5252	Sequence 5252, Ap
c 117	17.2	68.8	2372	3	US-09-814-353-21640	Sequence 21640, Ap
c 118	17.2	68.8	3888	7	US-10-437-963-21899	Sequence 21899, A
c 119	17.2	68.8	4407	9	US-10-480-988-45	Sequence 45, Ap1
c 120	17.2	68.8	5054	8	US-10-723-860-5201	Sequence 5201, Ap
c 121	17.2	68.8	33486	6	US-10-085-959-143	Sequence 143, App
c 122	17.2	68.8	35236	5	US-10-087-192-370	Sequence 370, App
c 123	17.2	68.8	47841	7	US-10-433-580-3	Sequence 3, App11
c 124	17.2	68.8	76180	7	US-10-322-281-492	Sequence 492, App
c 125	17.2	68.8	80959	3	US-09-858-546-3	Sequence 3, App11
c 126	17.2	68.8	143239	8	US-10-723-860-546	Sequence 546, App
c 127	17.2	68.8	143239	9	US-10-756-149-566	Sequence 566, App
c 128	17	68.0	136	3	US-09-815-242-2598	Sequence 2598, Ap
c 129	17	68.0	136	7	US-10-282-122A-5140	Sequence 5140, Ap
c 130	17	68.0	193	3	US-09-864-761-23226	Sequence 23226, A
c 131	17	68.0	229	7	US-10-424-599-6273	Sequence 6273, Ap
c 132	17	68.0	309	9	US-10-779-543-2758	Sequence 2758, Ap
c 133	17	68.0	349	3	US-09-770-791-721	Sequence 721, App
c 134	17	68.0	388	9	US-10-779-543-16188	Sequence 16188, A
c 135	17	68.0	396	4	US-09-925-065A-272052	Sequence 272052, A
c 136	17	68.0	399	3	US-09-815-242-1432	Sequence 1432, Ap
c 137	17	68.0	399	7	US-10-282-122A-3996	Sequence 3996, Ap
c 138	17	68.0	402	3	US-09-960-352-8750	Sequence 8750, Ap
c 139	17	68.0	402	7	US-10-437-963-4507	Sequence 4507, Ap
c 140	17	68.0	406	7	US-10-424-599-124554	Sequence 124554, A
c 141	17	68.0	456	3	US-09-864-761-6513	Sequence 6513, Ap
c 142	17	68.0	464	8	US-10-357-930-9911	Sequence 9911, Ap
c 143	17	68.0	477	7	US-10-430-201-1124	Sequence 1124, Ap
c 144	17	68.0	477	8	US-10-430-201-1125	Sequence 1125, Ap
c 145	17	68.0	482	8	US-10-425-115-162668	Sequence 162668, A
c 146	17	68.0	486	3	US-09-918-993-10324	Sequence 10324, A
c 147	17	68.0	489	3	US-09-815-242-4216	Sequence 4216, Ap
c 148	17	68.0	501	3	US-09-815-242-8952	Sequence 8352, Ap
c 149	17	68.0	501	7	US-10-282-122A-7624	Sequence 7624, Ap
c 150	17	68.0	501	8	US-10-857-625-233	Sequence 233, App

us-10-719-900-30

SEQ ID NO 30

LENGTH: 25

TYPE: DNA

ORGANISM: Mus musculus

US-10-719-900-30

Query Match

Best Local Similarity 100.0%; Score 25; DB 8; Length 25;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1

AAAAAATGGAATCTCAGGCTGAGA

25

Db

1

AAAAAATGGAATCTCAGGCTGAGA

25

RESULT 2

US-10-719-900-29

Sequence 29, Application US/10719900

Publication No. US20050026164A1

GENERAL INFORMATION:

APPLICANT: Xue Mei Zhou

TITLE OF INVENTION: Methods of Genetic Analysis of Mouse

FILE REFERENCE: 3528.1

CURRENT APPLICATION NUMBER: US/10/719,900

CURRENT FILING DATE: 2003-11-20

PRIOR APPLICATION NUMBER: 60/427,808

PRIOR FILING DATE: 2002 11 20

NUMBER OF SEQ ID NOS: 982914

SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

SEQ ID NO 29

LENGTH: 25

TYPE: DNA

ORGANISM: Mus musculus

US-10-719-900-29

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 8; Length 25;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1

AAAAAATGGAATCTCAGGCTGAGA

25

Db

1

AAAAAATGGAATCTCAGGCTGAGA

25

RESULT 3

US-09-925-065A-401109/c

Sequence 401109, Application US/09925065A

Publication No. US20050228172A9

GENERAL INFORMATION:

APPLICANT: Wang David G.

TITLE OF INVENTION: Identification and Mapping of Single

FILE REFERENCE: 108827.135

CURRENT APPLICATION NUMBER: US/09/925,065A

CURRENT FILING DATE: 2001-08-08

PRIOR APPLICATION NUMBER: US 60/243,096

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: US 60/252,147

PRIOR FILING DATE: 2000-11-20

PRIOR APPLICATION NUMBER: US 60/250,092

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: US 60/261,766

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/289,846

PRIOR FILING DATE: 2001-05-09

NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 401109

LENGTH: 583

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-065A-401109

ALIGNMENTS

RESULT 1  
US-10-719-900-30  
Sequence 30, Application US/10719900  
Publication No. US20050026164A1  
GENERAL INFORMATION:  
APPLICANT: Xue Mei Zhou  
TITLE OF INVENTION: Methods of Genetic Analysis of Mouse  
FILE REFERENCE: 3528.1  
CURRENT APPLICATION NUMBER: US/10/719,900  
CURRENT FILING DATE: 2003-11-20  
PRIOR APPLICATION NUMBER: 60/427,808  
PRIOR FILING DATE: 2002 11 20  
NUMBER OF SEQ ID NOS: 982914  
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1



```
QY 3 AAAAAGTGAATCTCAGGCTGAG 24
| | | | | | | | | | | | | | | |
Db 442 ACAACTGGAATCTCATGCTGAG 463

RESULT 7
US-10-741-601-5650/c
; Sequence 5650, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 471
; LENGTH: 2066
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-5650

Query Match 75.2%; Score 18.8; DB 7; Length 63693;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGAG 24
| | | | | | | | | | | | | | | |
Db 44895 ACAACTGGAATCTCATGCTGAG 44874

RESULT 8
US-09-925-065A-517136
; Sequence 517136, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 517136
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-517136

Query Match 74.4%; Score 18.6; DB 4; Length 589;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGTGAATCTCAGGCTGAGA 25
| | | | | | | | | | | | | | | |
Db 407 AAAGAAGTGAATCTCAGGCTGAGA 431

RESULT 9
US-10-094-749-471
; Sequence 471, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 471
; LENGTH: 2066
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-471

Query Match 74.4%; Score 18.6; DB 6; Length 2066;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGTGAATCTCAGGCTGAGA 25
| | | | | | | | | | | | | | | |
Db 820 AAAAATCGAAACTCGGGCTGAGA 844

RESULT 10
US-09-925-065A-417056
; Sequence 417056, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 417056
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-471
```

```
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(633)
; OTHER INFORMATION: n = A,T,C or G
; US-09-925-065A-417056

Query Match      73.6%; Score 18.4; DB 4; Length 633;
Best Local Similarity 95.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAATCGGAATCTCAGGCT 21
   ||||| ||||| ||||| |||||
Db 499 AAAATCTGGAATCTCAGGCT 518

RESULT 11
US-09-925-065A-637551/c
; Sequence 637551, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 637551
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-637551

Query Match      72.8%; Score 18.2; DB 4; Length 623;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAATCGGAATCTCAGGCTGAGA 25
   ||||| ||||| ||||| |||||
Db 615 AAAAATGTAGTCACAGGCTGAGA 593

RESULT 12
US-09-925-065A-523476/c
; Sequence 523476, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 523477
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-523477

Query Match      72.8%; Score 18.2; DB 4; Length 633;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAATCGGAATCTCAGGCTGAGA 25
   ||||| ||||| ||||| |||||
Db 49 AAAAATTTGGAATCTCAGGCTGAGA 27

RESULT 14
US-10-027-632-33065
; Sequence 33065, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
```



```
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33065
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-33065

Query Match          72.8%; Score 18.2; DB 5; Length 823;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGACA 25
   ||||| ||||| ||||| ||||| |||||
Db 257 AAAAACAATCTCAGGCTGACA 279

RESULT 15
US-10-027-632-33066
; Sequence 33066, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33066
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-33066

Query Match          72.8%; Score 18.2; DB 5; Length 823;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGACA 25
   ||||| ||||| ||||| ||||| |||||
Db 257 AAAAACAATCTCAGGCTGACA 279

RESULT 16
US-10-027-632-33065
; Sequence 33065, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33065
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-33065

Query Match          72.8%; Score 18.2; DB 6; Length 823;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGACA 25
   ||||| ||||| ||||| ||||| |||||
Db 257 AAAAACAATCTCAGGCTGACA 279

RESULT 17
US-10-027-632-33066
; Sequence 33066, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33066
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-33066

Query Match          72.8%; Score 18.2; DB 6; Length 823;
Best Local Similarity 87.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 3 AAAAAGTGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 257 AAAAATACATCTCAGGCTGACA 279

RESULT 18  
US-09-925-065A-362629  
; Sequence 362629, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 362629  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-362629

Query Match 71.2%; Score 17.8; DB 4; Length 525;  
Best Local Similarity 90.5%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTG 22  
||||| ||||| ||||| ||||| |||||  
Db 32 AAAAAGTGAATCTCAGGCTG 52

RESULT 19  
US-09-925-065A-362630  
; Sequence 362630, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 362630  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-362630

Query Match 71.2%; Score 17.8; DB 4; Length 525;

Best Local Similarity 90.5%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTG 22  
||||| ||||| ||||| ||||| |||||  
Db 32 AAAAAGTGAATCTCAGGCTG 52

RESULT 20  
US-09-925-065A-362631  
; Sequence 362631, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 362631  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-362631

Query Match 71.2%; Score 17.8; DB 4; Length 525;  
Best Local Similarity 90.5%; Pred. No. 3.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTG 22  
||||| ||||| ||||| ||||| |||||  
Db 32 AAAAAGTGAATCTCAGGCTG 52

RESULT 21  
US-10-027-632-202264/c  
; Sequence 202264, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 202264  
; LENGTH: 575  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-202264

Query Match 71.2%; Score 17.8; DB 5; Length 575;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTG 22  
||| ||||| ||||| ||||| |||||  
Db 157 AAGAAGTGAATCTCAGGCTG 137

RESULT 22  
US-10-027-632-202264/c  
; Sequence 202264, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; POLYMORPHISMS in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 202264  
; LENGTH: 575  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-202264

Query Match 71.2%; Score 17.8; DB 6; Length 575;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTG 22  
||| ||||| ||||| ||||| |||||  
Db 157 AAGAAGTGAATCTCAGGCTG 137

RESULT 23  
US-10-425-114-13369/c  
; Sequence 13369, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; THE PRODUCTION OF PLANTS FOR IMPROVED PROPERTIES  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 13369  
; LENGTH: 876  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: JC-GMST02400057G10\_FLI  
US-10-425-114-13369

Query Match 71.2%; Score 17.8; DB 7; Length 876;  
Best Local Similarity 90.5%; Pred. No. 3.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGAG 24  
||||| ||||| ||||| ||||| |||||  
Db 258 AAACCTGGAATCTCAGGCTGAG 238

RESULT 24  
US-10-424-599-113563/c  
; Sequence 113563, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; THE PRODUCTION OF PLANTS FOR IMPROVED PROPERTIES  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 113563  
; LENGTH: 884  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_73558C.1  
US-10-424-599-113563

Query Match 71.2%; Score 17.8; DB 7; Length 884;  
Best Local Similarity 90.5%; Pred. No. 3.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACCTGGAATCTCAGGCTGAG 24  
||||| ||||| ||||| ||||| |||||  
Db 266 AAACCTGGAATCTCAGGCTGAG 246

RESULT 25  
US-10-369-493-24074  
; Sequence 24074, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 24074  
; LENGTH: 1911  
; TYPE: DNA  
; ORGANISM: Xenorhabdus nematophilus  
US-10-369-493-24074

```
Query Match          71.2%; Score 17.8; DB 6; Length 1911;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAGTGGAAATCTCAGGCTGA 23
   |||||
Db 909 AAAAAGTGGAAATCTCTGGTTGA 929

RESULT 26
US-10-741-600-53948
; Sequence 53948, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53948
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-53948

Query Match          70.4%; Score 17.6; DB 8; Length 201;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGTGGAAATCTCAGGCTGAG 24
   |||||
Db 102 AAAAAGTGGAAATCTACAGGGTGAG 125

RESULT 27
US-09-933-797-614/c
; Sequence 614, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; Tissue
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/09/482,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 614
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Murine
US-09-933-797-614

Query Match          70.4%; Score 17.6; DB 3; Length 214;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAGTGGAAATCTCAGGCTGAGA 25
   |||||
Db 58 AATCACTGGAGGCTCAGGCTGAGA 35

RESULT 28
```

```
US-10-242-535A-41892/c
; Sequence 41892, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41892
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-41892

Query Match          70.4%; Score 17.6; DB 7; Length 383;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGTGGAAATCTCAGGCTGAG 24
   |||||
Db 141 AAAAGCACTGGAAACCAACGAGCTGAG 118

RESULT 29
US-10-085-783A-41892/c
; Sequence 41892, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41892
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-41892

Query Match          70.4%; Score 17.6; DB 7; Length 383;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGTGGAAATCTCAGGCTGAG 24
   |||||
Db 141 AAAAGCACTGGAAACCAACGAGCTGAG 118

RESULT 30
US-09-925-065A-311398
; Sequence 311398, Application US/09925065A
; Publication No. US20050228172A9
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311398
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-311398
```

```
Query Match 70.4%; Score 17.6; DB 4; Length 514;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
|||||
Db 247 AAAAGCTGGAGTCACAGGCTGAGA 270
```

## RESULT 31

```
US-09-925-065A-150366/c
; Sequence 150366, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150366
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-150366
```

```
Query Match 70.4%; Score 17.6; DB 4; Length 516;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTGGAATCTCAGGCTGAG 24
|||||
Db 215 AAAAACTGAAATGTAGGCTGTG 192
```

## RESULT 32

```
US-10-029-386-2155/c
; Sequence 2155, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2155
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR15.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
; OTHER INFORMATION: NT HIT: gi14756730, EVALUATE 1.00e-84
; OTHER INFORMATION: EST_HUMAN HIT: BE544267.1, EVALUATE 2.00e-84
US-10-029-386-2155
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 522;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTGGAATCTCAGGCTGAG 24
|||||
Db 264 AAAAACTGAAATGTAGGCTGTG 241
```

## RESULT 33

```
US-09-925-065A-501583
; Sequence 501583, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501583
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-501583
```

```
Query Match 70.4%; Score 17.6; DB 4; Length 538;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAACTGGAATCTCAGGCTGAG 24
|||||
```

Db 272 AAAAATCTGGAATTACAGGCTGAG 295

## RESULT 34

US-09-925-065A-435469/c  
; Sequence 435469, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435469  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-435469

Query Match 70.4%; Score 17.6; DB 4; Length 549;  
Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAATCTGGAATCTCAGGCTGAGA 25  
Db 417 AAAAATCTGGATCTCAGAAAGAGA 394

## RESULT 35

US-09-925-065A-435470/c  
; Sequence 435470, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435470  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-435470

Query Match 70.4%; Score 17.6; DB 4; Length 549;  
Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAATCTGGAATCTCAGGCTGAGA 25  
Db 417 AAAAATCTGGATCTCAGAAAGAGA 394

## RESULT 36

US-09-925-065A-435471/c  
; Sequence 435471, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435471  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-435471

Query Match 70.4%; Score 17.6; DB 4; Length 549;  
Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAATCTGGAATCTCAGGCTGAGA 25  
Db 417 AAAAATCTGGATCTCAGAAAGAGA 394

## RESULT 37

US-09-925-065A-435473/c  
; Sequence 435473, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435473  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-435473

```
Query Match 70.4%; Score 17.6; DB 4; Length 549;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||||| ||||| |||||
Db 417 AAAAACTGGGATCTCAGAAAGAGA 394

RESULT 38
US-11-060-756-455
; Sequence 455, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 455
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-455

Query Match 70.4%; Score 17.6; DB 10; Length 600;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||||| ||||| |||||
Db 211 AAAAACTAGAACTCTGAAGGTGAGA 234

RESULT 39
US-11-060-756-456
; Sequence 456, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 456
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-456

Query Match 70.4%; Score 17.6; DB 10; Length 600;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||||| ||||| |||||
Db 211 AAAAACTAGAACTCTGAAGGTGAGA 234

RESULT 40
US-11-060-756-4727
; Sequence 4727, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
```

```
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4727
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-4727

Query Match 70.4%; Score 17.6; DB 10; Length 600;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||||| ||||| |||||
Db 211 AAAAACTAGAACTCTGAAGGTGAGA 234

RESULT 41
US-11-060-756-4728
; Sequence 4728, Application US/11060756
; Publication No. US20050221354A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042000)
; CURRENT APPLICATION NUMBER: US/11/060,756
; CURRENT FILING DATE: 2005-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4728
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-060-756-4728

Query Match 70.4%; Score 17.6; DB 10; Length 600;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25
||||| ||||| ||||| |||||
Db 211 AAAAACTAGAACTCTGAAGGTGAGA 234

RESULT 42
US-09-925-065A-288955
; Sequence 288955, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
```



```
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288955
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-288955

Query Match          70.4%; Score 17.6; DB 4; Length 604;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| ||||| |||||
Db 329 AAAAACTGGAATCTCAGGCTGAG 352

RESULT 43
US-09-925-065A-265864/c
; Sequence 265864, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265864
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-265864

Query Match          70.4%; Score 17.6; DB 4; Length 606;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| ||||| |||||
Db 474 AAAGAGTGGCATCTCAGGCTAAG 451

RESULT 44
US-09-925-065A-563312
; Sequence 563312, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 563312
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-563312

Query Match          70.4%; Score 17.6; DB 4; Length 631;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| ||||| |||||
Db 323 AAAAACTGGAATCTCAGGCTGAG 346

RESULT 45
US-10-027-632-191799/c
; Sequence 191799, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 191799
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-191799

Query Match          70.4%; Score 17.6; DB 5; Length 656;
Best Local Similarity 83.3%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24
    ||||| ||||| ||||| ||||| |||||
Db 123 AAAAACTGCAATCTCATGCTGAG 100

RESULT 46
US-10-027-632-191799/c
; Sequence 191799, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 191799  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-191799

Query Match 70.4%; Score 17.6; DB 6; Length 656;  
Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
||| ||||| ||||| ||||| |||||  
DB 123 AAGAACTGCAATCTCATGCTGAG 100

RESULT 47  
US-09-925-065A-721165  
; Sequence 721165, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 721165  
; LENGTH: 1469  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-721165

Query Match 70.4%; Score 17.6; DB 4; Length 1469;  
Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
||| ||||| ||||| ||||| |||||  
DB 491 AAATATGGAATCTGAGGCTGAGA 514

RESULT 48  
US-10-424-599-123483/c  
; Sequence 123483, Application US/10424599  
; Publication No. US20040031072A1

; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 123483  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_82512C.1  
US-10-424-599-123483

Query Match 70.4%; Score 17.6; DB 7; Length 1551;  
Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
||| ||||| ||||| ||||| |||||  
DB 1325 AAAAACTGGAATTTGAGGCAAGA 1302

RESULT 49  
US-10-044-090-354/c  
; Sequence 354, Application US/10044090  
; Publication No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: PERL Program  
; SEQ ID NO 354  
; LENGTH: 5271  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020137081A1 221500.1  
; NAME/KEY: unsure  
; LOCATION: 1086, 1089, 1773-2110  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-044-090-354

Query Match 70.4%; Score 17.6; DB 5; Length 5271;  
Best Local Similarity 83.3%; Pred. No. 6.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
||| ||||| ||||| ||||| |||||  
DB 3345 AAAATACTGGGATTACAGGCTGAG 3322

RESULT 50  
US-10-084-817-340/c  
; Sequence 340, Application US/10084817  
; Publication No. US20030119009A1  
; GENERAL INFORMATION:  
; APPLICANT: Susan Stuart  
; APPLICANT: Jed G. Nuchtern  
; APPLICANT: Sharon E. Plon  
; APPLICANT: Jason M. Shohet  
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
; FILE REFERENCE: PA-0046 US  
; CURRENT APPLICATION NUMBER: US/10/084,817

Mon Feb 6 12:23:17 2006

; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/270,784  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 365  
; SOFTWARE: PERL Program  
; SEQ ID NO 340  
; LENGTH: 5271  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030119009A1 221500.1  
; NAME/KEY: unsure  
; LOCATION: 1086, 1089, 1773-2110  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-084-817-340

Query Match 70.4%; Score 17.6; DB 5; Length 5271;  
Best Local Similarity 83.3%; Pred. No. 6.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAACTGGATCTCAGGCTGAG 24  
||| ||||| ||| |||||  
Db 3345 AAAATACGGGATTACAGGCTGAG 3322

Search completed: February 3, 2006, 15:44:05  
Job time : 368.556 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-30

Perfect score: 25

Sequence: 1 aaaaacggaatcagctgaga 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US12\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US13\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US14\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US15\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US16\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	75.2	63693	7 US-10-995-561-13269	Sequence 13269, A
C 2	18.6	74.4	682	7 US-10-750-185-58129	Sequence 58129, A
C 3	18.6	74.4	682	7 US-10-750-623-58129	Sequence 58129, A
C 4	18.6	74.4	1231	7 US-10-750-185-46140	Sequence 46140, A
C 5	18.6	74.4	1231	7 US-10-750-623-46140	Sequence 46140, A
C 6	18.4	73.6	187745	8 US-11-121-086-83	Sequence 83, Appl
C 7	17.6	70.4	201	7 US-10-995-561-58459	Sequence 58459, A
C 8	17.6	70.4	835	7 US-10-750-185-33984	Sequence 33984, A
C 9	17.6	70.4	835	7 US-10-750-623-33984	Sequence 33984, A
C 10	17.6	70.4	51917	7 US-10-995-561-13338	Sequence 63, Appl
C 11	17.4	69.6	169725	8 US-11-121-086-63	Sequence 1656, Ap
C 12	17.2	68.8	479	7 US-10-750-185-1656	Sequence 1656, Ap
C 13	17.2	68.8	479	7 US-10-750-623-1656	Sequence 1656, Ap
C 14	17.2	68.8	750	7 US-10-750-185-37889	Sequence 37889, A
C 15	17.2	68.8	750	7 US-10-750-623-37889	Sequence 37889, A
C 16	17.2	68.8	1768	7 US-10-750-185-32085	Sequence 32085, A
C 17	17.2	68.8	1768	7 US-10-750-623-32085	Sequence 32085, A
C 18	17.2	68.0	463	8 US-11-000-688-898	Sequence 898, App
C 19	17.2	68.0	463	8 US-11-000-688-898	Sequence 898, App
C 20	17.2	68.0	475	7 US-10-750-185-1913	Sequence 1913, Ap
C 21	17.2	68.0	600	7 US-10-750-623-1913	Sequence 1913, Ap
C 22	17.2	68.0	1094	8 US-11-136-527-285	Sequence 285, App

```
c 96 16 64.0 992 7 US-10-750-623-53909 Sequence 53909, A
c 97 16 64.0 1408 7 US-10-750-185-44804 Sequence 44804, A
c 98 16 64.0 1408 7 US-10-750-623-44804 Sequence 44804, A
c 99 16 64.0 1438 7 US-10-750-185-39949 Sequence 39949, A
c 100 16 64.0 1438 7 US-10-750-623-39949 Sequence 39949, A
c 101 16 64.0 1550 8 US-11-000-688-323 Sequence 323, App
c 102 16 64.0 1601 7 US-10-750-185-27423 Sequence 27423, A
c 103 16 64.0 1601 7 US-10-750-623-27423 Sequence 27423, A
c 104 16 64.0 1679 7 US-10-750-185-28409 Sequence 28409, A
c 105 16 64.0 1679 7 US-10-750-623-28409 Sequence 28409, A
c 106 16 64.0 1695 7 US-10-750-185-55993 Sequence 55993, A
c 107 16 64.0 1695 7 US-10-750-623-55993 Sequence 55993, A
c 108 16 64.0 1697 7 US-10-750-185-42862 Sequence 42862, A
c 109 16 64.0 1697 7 US-10-750-623-42862 Sequence 42862, A
c 110 16 64.0 1828 7 US-10-750-185-55873 Sequence 55873, A
c 111 16 64.0 1828 7 US-10-750-623-55873 Sequence 55873, A
c 112 16 64.0 1851 7 US-10-750-185-62541 Sequence 62541, A
c 113 16 64.0 1851 7 US-10-750-623-62541 Sequence 62541, A
c 114 16 64.0 1993 7 US-10-750-185-54772 Sequence 54772, A
c 115 16 64.0 1993 7 US-10-750-623-54772 Sequence 54772, A
c 116 16 64.0 2001 8 US-11-043-752-3895 Sequence 3895, App
c 117 16 64.0 2001 8 US-11-043-752-3896 Sequence 3896, App
c 118 16 64.0 2061 8 US-11-128-061-3438 Sequence 3438, App
c 119 16 64.0 2061 8 US-11-128-049-3438 Sequence 3438, App
c 120 16 64.0 2269 8 US-11-000-688-1374 Sequence 1374, App
c 121 16 64.0 2305 7 US-10-750-185-30133 Sequence 30133, A
c 122 16 64.0 2305 7 US-10-750-623-30133 Sequence 30133, A
c 123 16 64.0 2489 7 US-10-750-185-53608 Sequence 53608, A
c 124 16 64.0 2489 7 US-10-750-623-53608 Sequence 53608, A
c 125 16 64.0 2562 7 US-10-750-185-54101 Sequence 54101, A
c 126 16 64.0 2562 7 US-10-750-623-54101 Sequence 54101, A
c 127 16 64.0 2576 7 US-10-750-185-48750 Sequence 48750, A
c 128 16 64.0 2576 7 US-10-750-623-48750 Sequence 48750, A
c 129 16 64.0 2608 7 US-10-750-185-57741 Sequence 57741, A
c 130 16 64.0 2608 7 US-10-750-623-57741 Sequence 57741, A
c 131 16 64.0 2838 7 US-10-750-185-46440 Sequence 46440, A
c 132 16 64.0 2838 7 US-10-750-623-46440 Sequence 46440, A
c 133 16 64.0 2846 7 US-10-750-185-38714 Sequence 38714, A
c 134 16 64.0 2846 7 US-10-750-623-38714 Sequence 38714, A
c 135 16 64.0 3334 7 US-10-750-185-54572 Sequence 54572, A
c 136 16 64.0 3334 7 US-10-750-623-54572 Sequence 54572, A
c 137 16 64.0 3526 7 US-10-750-185-25056 Sequence 25056, A
c 138 16 64.0 3526 7 US-10-750-623-25056 Sequence 25056, A
c 139 16 64.0 3769 7 US-10-750-185-25076 Sequence 25076, A
c 140 16 64.0 3769 7 US-10-750-623-25076 Sequence 25076, A
c 141 16 64.0 3769 7 US-10-750-185-38543 Sequence 38543, A
c 142 16 64.0 3769 7 US-10-750-623-38543 Sequence 38543, A
c 143 16 64.0 3827 8 US-11-136-527-3008 Sequence 3008, App
c 144 16 64.0 5039 7 US-10-750-185-31557 Sequence 31557, A
c 145 16 64.0 5039 7 US-10-750-623-31557 Sequence 31557, A
c 146 16 64.0 18097 7 US-10-995-561-13486 Sequence 13486, A
c 147 16 64.0 18097 7 US-10-995-561-13486 Sequence 13486, A
c 148 16 64.0 20774 7 US-10-995-561-13291 Sequence 13291, A
c 149 16 64.0 23606 7 US-10-995-561-13465 Sequence 13465, A
c 150 16 64.0 96109 8 US-11-124-368A-2895 Sequence 2895, App
```

ALIGNMENTS

```
RESULT 1
US-10-995-561-13269/c
; Sequence 13269, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702

Query Match 75.2%; Score 18.8; DB 7; Length 63693;
Best Local Similarity 90.9%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAGTGAATCTCAGGCTGAG 24
| | | | | | | | | | | | | | | | | | | | | |
Db 44895 ACAACTGGAATCTCATGCTGAG 44874
| | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-10-750-185-58129
; Sequence 58129, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58129
; LENGTH: 682
; TYPE: DNA
; ORGANISM: Bovine 19866881916918
US-10-750-185-58129

Query Match 74.4%; Score 18.6; DB 7; Length 682;
Best Local Similarity 84.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGTGAATCTCAGGCTGAGA 25
| | | | | | | | | | | | | | | | | | | | | |
Db 110 AAAAAAGTGAATCTCTAGCTGAGA 134
| | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-10-750-623-58129
; Sequence 58129, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58129
; LENGTH: 682
```





;  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33984  
; LENGTH: 835  
; TYPE: DNA  
; ORGANISM: Bovine 19866881966547  
US-10-750-185-33984

Query Match 70.4%; Score 17.6; DB 7; Length 835;  
Best Local Similarity 83.3%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||||| ||||||| |||||||  
Db 121 AAAAAATGGAATCTCCCGATGAGA 98

RESULT 9  
US-10-750-623-33984/c  
; Sequence 33984, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33984  
; LENGTH: 835  
; TYPE: DNA  
; ORGANISM: Bovine 19866881966547  
US-10-750-623-33984

Query Match 70.4%; Score 17.6; DB 7; Length 835;  
Best Local Similarity 83.3%; Pred. No. 69;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||||| ||||||| |||||||  
Db 121 AAAAAATGGAATCTCCCGATGAGA 98

RESULT 10  
US-10-995-561-13338  
; Sequence 13338, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

;  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13338  
; LENGTH: 51917  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)....(51917)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-10-995-561-13338

Query Match 70.4%; Score 17.6; DB 7; Length 51917;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAG 24  
||||| ||||||| ||||||| |||||||  
Db 27042 AAAATACTGGAATTACAGGCTGAG 27065

RESULT 11  
US-11-121-086-63/c  
; Sequence 63, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 63  
; LENGTH: 169725  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-63

Query Match 69.6%; Score 17.4; DB 8; Length 169725;  
Best Local Similarity 94.7%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 ACTGGAATCTCAGGCTGAG 24  
||||| ||||||| ||||||| |||||||  
Db 110537 ACTGTAATCTCAGGCTGAG 110519

RESULT 12  
US-10-750-185-1656/c  
; Sequence 1656, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 1656  
; LENGTH: 479  
; TYPE: DNA  
; ORGANISM: Bovine MMBT20262  
US-10-750-185-1656

Query Match 68.8%; Score 17.2; DB 7; Length 479;  
Best Local Similarity 86.4%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGA 23  
|||||  
DB 177 AAAAAGCACTCTCAGGCTGA 156

RESULT 13  
US-10-750-623-1656/c  
; Sequence 1656, Application US/10750623  
; Publication No. US20050287531A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 1656  
; LENGTH: 479  
; TYPE: DNA  
; ORGANISM: Bovine MMBT20262  
US-10-750-623-1656

Query Match 68.8%; Score 17.2; DB 7; Length 479;  
Best Local Similarity 86.4%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGA 23  
|||||  
DB 177 AAAAAGCACTCTCAGGCTGA 156

RESULT 14  
US-10-750-185-37889/c  
; Sequence 37889, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1

; SEQ ID NO 37889  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Bovine 19866881213939  
US-10-750-185-37889

Query Match 68.8%; Score 17.2; DB 7; Length 750;  
Best Local Similarity 86.4%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGA 23  
|||||  
DB 448 AAAAAGCACTCTCAGGCTGA 427

RESULT 15  
US-10-750-623-37889/c  
; Sequence 37889, Application US/10750623  
; Publication No. US20050287531A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 37889  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Bovine 19866881213939  
US-10-750-623-37889

Query Match 68.8%; Score 17.2; DB 7; Length 750;  
Best Local Similarity 86.4%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGTGAATCTCAGGCTGA 23  
|||||  
DB 448 AAAAAGCACTCTCAGGCTGA 427

RESULT 16  
US-10-750-185-32085/c  
; Sequence 32085, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 32085  
; LENGTH: 1768  
; TYPE: DNA

```
; ORGANISM: Bovine 19866880282614
US-10-750-185-32085

Query Match      68.8%; Score 17.2; DB 7; Length 1768;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTGGAATCTCAGGCTGA 23
      || ||||| ||||| ||||| |||||
Db      1464 AATGACTGGACTCTCAGGCTGA 1443

RESULT 17
US-10-750-623-32085/c
; Sequence 32085, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32085
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Bovine 19866880282614
US-10-750-623-32085

Query Match      68.8%; Score 17.2; DB 7; Length 1768;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTGGAATCTCAGGCTGA 23
      || ||||| ||||| ||||| |||||
Db      1464 AATGACTGGACTCTCAGGCTGA 1443

RESULT 18
US-11-000-688-898/c
; Sequence 898, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 898
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; NAME/KEY: misc_feature
; LOCATION: (1)..(463)
; OTHER INFORMATION: 3' terminal sequence from clone
; OTHER INFORMATION: 3' terminal sequence from clone

; ORGANISM: Bovine 19866880282614
US-10-750-185-32085

Query Match      68.8%; Score 17.2; DB 7; Length 1768;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAACTGGAATCTCAGGCTGA 23
      || ||||| ||||| ||||| |||||
Db      1464 AATGACTGGACTCTCAGGCTGA 1443

RESULT 19
US-11-000-688-899
; Sequence 899, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 899
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; NAME/KEY: misc_feature
; LOCATION: (1)..(475)
; OTHER INFORMATION: 5' terminal sequence from clone
; OTHER INFORMATION: image:358683.tropomyosin 1 (alpha) (TPM1) gene.
US-11-000-688-899

Query Match      68.0%; Score 17; DB 8; Length 475;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAACTGGAATCTCAGGCTGAGA 25
      || ||||| ||||| ||||| |||||
Db      439 AAGTCACTGGAGGCTCAGGCTGAGA 435

RESULT 20
US-10-750-185-1913
; Sequence 1913, Application US/10750185
; Publication No. US20050280603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
```

```

; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 1913
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT18521
US-10-750-185-1913

```

Query Match	68.0%	Score 17	DB 7	Length 600
Best Local Similarity	80.0%	Pred. No. 1.2e+02		
Matches	20	Conservative	0	Mismatches 5; Indels
Qy	1	AAAAAACTGGAAATCTCAGGCTGAGA	25	
Db	189	AACATACACTGGAAATCTCTGGGCTGAGA	213	

```

RESULT 21
US-10-750-623-1913
; Sequence 1913, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1913
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT18521
US-10-750-623-1913

```

```

Query Match      68.0%; Score 17; DB 7; Length 600;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTGAATCTCAGGCTGAGA 25
Db 189 AACATACTGGAATCTCTGGCTGAGA 213

```

```

RESULT 22
US-11-136-527-285
; Sequence 285, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 285
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-285

```

Query Match 68.0%; Score 17; DB 8; Length 1094;

```

Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0
Qy 1 AAAAACTGCAATCTCAGGCTGAGA 25
Db 641 AAGTCACTGGAGGCTCAGGCTGAGA 665

```

```

RESULT 23
US-11-136-527-4381
; Sequence 4381, Application US/11136527
; Publication NO. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4381
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-4381

```

```

Query Match      68.0%; Score 17; DB 8; Length 1094;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAACTGGAATCTCAGGCTGAGA 25
    |||||
Db 641 AAGTCACTGGAGGCTCAGGCTGAGA 665

```

```

RESULT 24
US-10-750-185-40453/c
; Sequence 40453, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 40453
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-40453

```

```
Query Match      68.0%; Score 17; DB 7; Length 1412;
Best Local Similarity 80.0%; Pred.No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels

QY    1 AAAAAACTGGAACTCTCAGGCTGAGA 25
      |||||
nB   363 AAAGACCTTAGAACAATTAGGCTTGAGA 339
```

```
RESULT 25
US-10-750-623-40453/c
; Sequence 40453, Application US/10750623
; Publication NO. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: PM1006290
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40453
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Bovine 19866880538167
US-10-750-623-40453

Query Match      68.0%; Score 17; DB 7; Length 1412;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 363 AAGACCTAGAAACTTAGGCTGAGA 339

RESULT 26
US-10-485-517-61
; Sequence 61, Application US/10485517
; Publication NO. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: BIOSYNEXUS INCORPORATED
; APPLICANT: FOSTER, SIMON
; APPLICANT: MOND, JAMES
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P1006290
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 2084
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-485-517-61

Query Match      68.0%; Score 17; DB 7; Length 2084;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 396 AACATCCGGAATCTTAATGCTGAGA 420

RESULT 27
US-10-485-517-76
; Sequence 76, Application US/10485517
; Publication NO. US20050256299A1
```

```
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: BIOSYNEXUS INCORPORATED
; APPLICANT: FOSTER, SIMON
; APPLICANT: MOND, JAMES
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P1006290
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76
; LENGTH: 2084
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-485-517-76

Query Match      68.0%; Score 17; DB 7; Length 2084;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 396 AACATCCGGAATCTTAATGCTGAGA 420

RESULT 28
US-10-750-185-31367
; Sequence 31367, Application US/10750185
; Publication NO. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31367
; LENGTH: 3272
; TYPE: DNA
; ORGANISM: Bovine 19866881031011
US-10-750-185-31367

Query Match      68.0%; Score 17; DB 7; Length 3272;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
Db 1804 AAAATAGGAATCTGAGGCTGAGA 1828

RESULT 29
US-10-750-623-31367
; Sequence 31367, Application US/10750623
; Publication NO. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
```

```
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31367
; LENGTH: 3272
; TYPE: DNA
; ORGANISM: Bovine 19866881031011
US-10-750-623-31367

Query Match      68.0%; Score 17; DB 7; Length 3272;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
    ||||| ||||| ||||| |||||
DB 1804 AAAAAAGGAATCTCAGGCTCAGA 1828

RESULT 30
US-10-821-234-62/c
; Sequence 62, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: Pt_SEQ_genes Version 1.0
; SEQ ID NO 62
; LENGTH: 5796
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-62

Query Match      68.0%; Score 17; DB 7; Length 5796;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAACTGGAATCTCAGGCTGAGA 25
    ||||| ||||| ||||| |||||
DB 3003 AAGAAGTGGACCTTGAGTTGAGA 2979

RESULT 31
US-10-310-914A-626925
; Sequence 626925, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiler, Kvazat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
```

```
; SEQ ID NO 626925
; LENGTH: 21
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-626925

Query Match      67.2%; Score 16.8; DB 7; Length 21;
Best Local Similarity 70.0%; Pred. No. 78;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACGTGAATCTCAGGCTGA 23
    ||| ||||| :||| |||||
DB 2 AACCCUGGACUCUCAGGCUGA 21

RESULT 32
US-10-310-914A-626926
; Sequence 626926, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiler, Kvazat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 626926
; LENGTH: 23
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-626926

Query Match      67.2%; Score 16.8; DB 7; Length 23;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACGTGAATCTCAGGCTGA 23
    ||| ||||| :||| |||||
DB 2 AACCCUGGACUCUCAGGCUGA 21

RESULT 33
US-11-121-849-207753/c
; Sequence 207753, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; TITLE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 207753
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-207753

Query Match      67.2%; Score 16.8; DB 8; Length 25;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAACGTGAATCTCAGGCTGA 23
    ||| ||||| ||||| |||||
DB 21 AACCTGGACTCTCAGGCTGA 2
```

```
RESULT 34
US-11-136-527-207659
; Sequence 207659, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 207659
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-207659

Query Match      67.2%; Score 16.8; DB 8; Length 25;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      6 ACTGGAATCTCAGGCTGAGA 25
DB      3 ACTGGAGGCTCAGGCTGAGA 22

RESULT 35
US-10-750-185-27917/c
; Sequence 27917, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27917
; LENGTH: 860
; TYPE: DNA
; ORGANISM: Bovine 19866881073914
US-10-750-185-27917

Query Match      67.2%; Score 16.8; DB 7; Length 860;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAAACTGGAATCTCAGGCT 21
DB      603 AAAAATGGCATCTCAGGTT 584

RESULT 36
US-10-750-623-27917/c
; Sequence 27917, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
```

```
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27917
; LENGTH: 860
; TYPE: DNA
; ORGANISM: Bovine 19866881073914
US-10-750-623-27917

Query Match      67.2%; Score 16.8; DB 7; Length 860;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAAACTGGAATCTCAGGCT 21
DB      603 AAAAATGGCATCTCAGGTT 584

RESULT 37
US-10-509-773-11/c
; Sequence 11, Application US/10509773
; Publication No. US20050281743A1
; GENERAL INFORMATION:
; APPLICANT: Delaney, Allen
; TITLE OF INVENTION: Cancer Associated Protein Phosphatases and their
; FILE REFERENCE: SMAR-044
; CURRENT APPLICATION NUMBER: US/10/509,773
; CURRENT FILING DATE: 2004-09-28
; PRIOR APPLICATION NUMBER: CA03/00393
; PRIOR FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 60/368,859
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: FLJ20313 nucleotide sequence
US-10-509-773-11

Query Match      67.2%; Score 16.8; DB 7; Length 2226;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 AAACCTGGAATCTCAGGCTGA 23
DB      2057 AACCTGGACTCTCAGGCTGA 2038

RESULT 38
US-10-750-185-2152/c
; Sequence 2152, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
```



; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 2152  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT21073  
US-10-750-185-2152

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 202 AAAACTGCAATCTCATCTTGAGA 180

RESULT 39  
US-10-750-185-3019  
; Sequence 3019, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 3019  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT21143  
US-10-750-185-3019

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 91 AGAACAGGCATCTCAGGCTCAGA 113

RESULT 40  
US-10-750-623-2152/c  
; Sequence 2152, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 2152  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT21073  
US-10-750-623-2152

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 202 AAAACTGCAATCTCATCTTGAGA 180

RESULT 41  
US-10-750-623-3019  
; Sequence 3019, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 3019  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT21143  
US-10-750-623-3019

Query Match 66.4%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAACTGGAATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 91 AGAACAGGCATCTCAGGCTCAGA 113

RESULT 42  
US-10-750-185-38194  
; Sequence 38194, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

FILE REFERENCE: MM1100-2  
CURRENT APPLICATION NUMBER: US/10/750,185  
CURRENT FILING DATE: 2003-12-31  
PRIOR APPLICATION NUMBER: US 60/437,482  
PRIOR FILING DATE: 2002-12-31  
NUMBER OF SEQ ID NOS: 64922  
SOFTWARE: PatentIN version 3.1  
SEQ ID NO 38194  
LENGTH: 1130  
TYPE: DNA  
ORGANISM: Bovine 1986680591492  
US-10-750-185-38194

Query Match 66.4%; Score 16.6; DB 7; Length 1130;  
Best Local Similarity 82.6%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAGGATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 541 AGACAGGCATCTCAGGCTCAGA 563

RESULT 43  
US-10-750-623-38194  
Sequence 38194, Application US/10750623  
Publication No. US20050287531A1  
GENERAL INFORMATION:  
APPLICANT: MMI GENOMICS, INC.  
APPLICANT: DENISE, Sue K.  
APPLICANT: KERR, Richard  
APPLICANT: ROSENFELD, David  
APPLICANT: HOLM, Tom  
APPLICANT: BATES, Stephen  
APPLICANT: FANTIN, Dennis  
TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
FILE REFERENCE: MM1100-1  
CURRENT APPLICATION NUMBER: US/10/750,623  
CURRENT FILING DATE: 2003-12-31  
PRIOR APPLICATION NUMBER: US 60/437,482  
PRIOR FILING DATE: 2002-12-31  
NUMBER OF SEQ ID NOS: 64922  
SOFTWARE: PatentIN version 3.1  
SEQ ID NO 38194  
LENGTH: 1130  
TYPE: DNA  
ORGANISM: Bovine 1986680591492  
US-10-750-623-38194

Query Match 66.4%; Score 16.6; DB 7; Length 1130;  
Best Local Similarity 82.6%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAGGATCTCAGGCTGAGA 25  
||||| ||||| ||||| ||||| |||||  
Db 541 AGACAGGCATCTCAGGCTCAGA 563

RESULT 44  
US-10-131-826A-259  
Sequence 259, Application US/10131826A  
Publication No. US20050245730A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME  
FILE REFERENCE: P3330R1C128  
CURRENT APPLICATION NUMBER: US/10/131,826A  
CURRENT FILING DATE: 2002-04-24  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 259  
LENGTH: 2024  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-131-826A-259

Query Match 66.4%; Score 16.6; DB 7; Length 2024;  
Best Local Similarity 82.6%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGTGGATCTCAGGCTGA 23  
||||| ||||| ||||| ||||| |||||  
Db 1419 AACAAAGTGGATTCATCTCGA 1441

RESULT 45  
US-10-750-185-37312/c  
Sequence 37312, Application US/10750185  
Publication No. US20050260603A1  
GENERAL INFORMATION:  
APPLICANT: MMI GENOMICS, INC.  
APPLICANT: DENISE, Sue K.  
APPLICANT: KERR, Richard  
APPLICANT: ROSENFELD, David  
APPLICANT: HOLM, Tom  
APPLICANT: BATES, Stephen  
APPLICANT: FANTIN, Dennis  
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
FILE REFERENCE: MM1100-2  
CURRENT APPLICATION NUMBER: US/10/750,185  
CURRENT FILING DATE: 2003-12-31  
PRIOR APPLICATION NUMBER: US 60/437,482  
PRIOR FILING DATE: 2002-12-31  
NUMBER OF SEQ ID NOS: 64922  
SOFTWARE: PatentIN version 3.1  
SEQ ID NO 37312  
LENGTH: 2078  
TYPE: DNA  
ORGANISM: Bovine 1986680966268  
US-10-750-185-37312



```

US-11-121-086-46
; Sequence 46, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 46
; LENGTH: 158410
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-46

Query Match      66.4%; Score 16.6; DB 8; Length 158410;
Best Local Similarity 82.6%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 AAAACTGGAATCTCAGGCTGAGA 25
Db      148716 AAATCTGGAATCCAGCGGAGA 148738

Search completed: February 3, 2006, 16:20:21
Job time : 347.111 secs

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:23 ; Search time 863 Seconds  
(without alignments)  
1646.682 Million cell updates/sec

Title: US-10-719-900-35  
Perfect score: 25  
Sequence: 1 aaaaagagcgagactggtcttc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_in:\*
- 3: gb\_env:\*
- 4: gb\_on:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pr:\*
- 9: gb\_ro:\*
- 10: gb\_sts:\*
- 11: gb\_sy:\*
- 12: gb\_un:\*
- 13: gb\_vi:\*
- 14: gb\_htg:\*
- 15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	163878	9 AC134563	AC134563 Mus muscu
2	23.4	93.6	193811	9 AC142098	AC142098 Mus muscu
3	21.8	87.2	28365	14 AC166153	AC166153 Mus muscu
4	21.8	87.2	98311	9 CR788287	CR788287 Mouse DNA
5	21.8	87.2	110000	14 BX546458_1	Continuation (2 of BX546458_1)
6	21.8	87.2	110000	14 BX546462_0	Continuation (3 of BX546462_0)
7	21.8	87.2	110000	14 CR753898_0	Continuation (3 of CR753898_0)
8	21.8	87.2	110000	14 CR753898_2	Continuation (3 of CR753898_2)
9	21.8	87.2	115602	9 BX284639	BX284639 Mouse DNA
10	21.8	87.2	136564	14 CR974476	CR974476 Mus muscu
11	21.8	87.2	143116	14 AC118310	AC118310 Rattus no
12	21.8	87.2	143176	14 AC166152	AC166152 Mus muscu
13	21.8	87.2	145110	9 CR354442	CR354442 Mouse DNA
14	21.8	87.2	145110	9 CR354442	CR354442 Mouse DNA
15	21.8	87.2	145376	14 CT010482	CT010482 Mus muscu
16	21.8	87.2	149369	9 AL607123	AL607123 Mouse DNA
17	21.8	87.2	151693	14 CR955032	CR955032 Homo sapi
18	21.8	87.2	169581	9 BX511235	BX511235 Mouse DNA

19	21.8	87.2	174479	14 CR955031	CR955031 Homo sapi
20	21.8	87.2	174479	14 CR955031	CR955031 Homo sapi
21	21.8	87.2	179228	9 BX005149	BX005149 Mouse DNA
22	21.8	87.2	179228	9 BX005149	BX005149 Mouse DNA
23	21.8	87.2	180947	9 AL845456	AL845456 Mouse DNA
24	21.8	87.2	180947	9 AL845456	AL845456 Mouse DNA
25	21.8	87.2	186050	9 BX682537	BX682537 Mouse DNA
26	21.8	87.2	190928	14 AC118842	AC118842 Rattus no
27	21.8	87.2	193326	9 CR318639	CR318639 Mouse DNA
28	21.8	87.2	198262	14 BX465218	BX465218 Mus muscu
29	21.8	87.2	200474	14 CR956625	CR956625 Mus muscu
30	21.8	87.2	201381	9 AL845491	AL845491 Mouse DNA
31	21.8	87.2	203209	14 AC165368	AC165368 Mus muscu
32	21.8	87.2	205459	14 CR848808	CR848808 Mouse DNA
33	21.8	87.2	205573	9 AL591032	AL591032 Mouse DNA
34	21.8	87.2	206224	14 AC027767	AC027767 Mus muscu
35	21.8	87.2	207814	9 AL590994	AL590994 Mouse DNA
36	21.8	87.2	208910	14 AC079573	AC079573 Mus muscu
37	21.8	87.2	213321	14 AC102988	AC102988 Rattus no
38	21.8	87.2	216517	9 AL845494	AL845494 Mouse DNA
39	21.8	87.2	217197	9 BX813319	BX813319 Mouse DNA
40	21.8	87.2	217867	14 AC095149	AC095149 Rattus no
41	21.8	87.2	218484	9 AL845476	AL845476 Mouse DNA
42	21.8	87.2	223248	14 AC055778	AC055778 Mus muscu
43	21.8	87.2	224765	14 AC106363	AC106363 Rattus no
44	21.8	87.2	226542	9 BX890623	BX890623 Mouse DNA
45	21.8	87.2	229093	14 AC128603	AC128603 Rattus no
46	21.8	87.2	235381	14 AC123015	AC123015 Rattus no
47	21.8	87.2	237321	14 AC103495	AC103495 Rattus no
48	21.8	87.2	238355	14 AC126306	AC126306 Rattus no
49	21.8	87.2	239591	14 AC128432	AC128432 Rattus no
50	21.8	87.2	239946	14 AC103505	AC103505 Rattus no
51	21.8	87.2	241531	14 AC122959	AC122959 Rattus no
52	21.8	87.2	241795	14 AC123410	AC123410 Rattus no
53	21.8	87.2	242915	14 AC117981	AC117981 Rattus no
54	21.8	87.2	245651	14 AC126537	AC126537 Rattus no
55	21.8	87.2	245993	14 AC130970	AC130970 Rattus no
56	21.8	87.2	248568	14 AC115184	AC115184 Rattus no
57	21.8	87.2	251251	14 AC130500	AC130500 Rattus no
58	21.8	87.2	255563	14 AC125834	AC125834 Rattus no
59	21.8	87.2	259702	14 AC119031	AC119031 Rattus no
60	21.8	87.2	274751	14 BX649574	BX649574 Mus muscu
61	21.8	87.2	278876	14 AC120936	AC120936 Rattus no
62	20.8	83.2	225269	14 AC133757	AC133757 Rattus no
63	20.8	83.2	244105	14 AC098897	AC098897 Rattus no
64	20.2	80.8	878	9 BC017151	BC017151 Mus muscu
65	20.2	80.8	7279	15 AK120499	AK120499 Oryza sat
66	20.2	80.8	33961	14 AC128246_5	Continuation (6 of AC128246_5)
67	20.2	80.8	34046	14 AC094670_5	Continuation (6 of AC094670_5)
68	20.2	80.8	40774	14 AC100658	AC100658 Mus muscu
69	20.2	80.8	47993	14 AC141397	AC141397 Rattus no
70	20.2	80.8	57402	14 AC127888_3	Continuation (4 of AC127888_3)
71	20.2	80.8	64074	14 AC117718	AC117718 Mus muscu
72	20.2	80.8	65102	14 AC120436	AC120436 Mus muscu
73	20.2	80.8	84257	14 AC099430_3	Continuation (4 of AC099430_3)
74	20.2	80.8	106784	14 AC136250	AC136250 Rattus no
75	20.2	80.8	106784	14 AC136250	AC136250 Rattus no
76	20.2	80.8	110000	14 AC091229_10	Continuation (11 of AC091229_10)
77	20.2	80.8	110000	14 AC091242_3	Continuation (4 of AC091242_3)
78	20.2	80.8	110000	14 AC091347_1	Continuation (2 of AC091347_1)
79	20.2	80.8	110000	14 AC091360_0	AC091360 Rattus no
80	20.2	80.8	110000	14 AC094469_1	Continuation (2 of AC094469_1)
81	20.2	80.8	110000	14 AC094670_4	Continuation (5 of AC094670_4)
82	20.2	80.8	110000	14 AC097996_4	Continuation (5 of AC097996_4)
83	20.2	80.8	110000	14 AC106346_1	Continuation (2 of AC106346_1)
84	20.2	80.8	110000	14 AC106541_3	Continuation (4 of AC106541_3)
85	20.2	80.8	110000	14 AC113631_0	AC113631 Rattus no
86	20.2	80.8	110000	14 AC114446_3	Continuation (4 of AC114446_3)
87	20.2	80.8	110000	14 AC120731_0	AC120731 Rattus no
88	20.2	80.8	110000	14 AC125991_3	Continuation (4 of AC125991_3)
89	20.2	80.8	110000	14 AC127888_2	Continuation (3 of AC127888_2)
90	20.2	80.8	110000	14 AC128246_4	Continuation (5 of AC128246_4)
91	20.2	80.8	110000	15 AP008215_049	Continuation (50 of AP008215_049)

Mon Feb 6 12:23:18 2006

```
92 20.2 80.8 119802 14 AC141020 Rattus no
93 20.2 80.8 125863 14 AC125505 Rattus no
94 20.2 80.8 130254 14 AC136195 Rattus no
c 95 20.2 80.8 141458 14 AC142044 Rattus no
c 96 20.2 80.8 142422 9 AL844588 Mouse DNA
c 97 20.2 80.8 143387 14 AC162885 Mus muscu
c 98 20.2 80.8 146166 14 AC147786 Rattus no
c 99 20.2 80.8 146508 14 AC114719 Rattus no
c 100 20.2 80.8 147258 14 AC141400 Rattus no
101 20.2 80.8 150849 14 AC141935 Rattus no
102 20.2 80.8 150936 9 AC087722 Rattus no
c 103 20.2 80.8 152788 14 AC120512 Rattus no
c 104 20.2 80.8 153819 9 AC121955 Mus muscu
c 105 20.2 80.8 155094 14 AC122080 Rattus no
c 106 20.2 80.8 157071 9 AC111718 Rattus no
c 107 20.2 80.8 157990 14 AC137311 Rattus no
c 108 20.2 80.8 158089 9 AL645852 Mouse DNA
c 109 20.2 80.8 160132 14 AC141371 Rattus no
c 110 20.2 80.8 160348 9 AC156275 Mus muscu
c 111 20.2 80.8 160537 9 AC129311 Mus muscu
c 112 20.2 80.8 160901 14 AC120060 Rattus no
c 113 20.2 80.8 161265 9 AL935320 Mouse DNA
c 114 20.2 80.8 163263 14 AC121352 Rattus no
c 115 20.2 80.8 164210 14 AC135478 Rattus no
c 116 20.2 80.8 164571 9 AL929033 Mouse DNA
c 117 20.2 80.8 164926 14 AC124922 Rattus no
c 118 20.2 80.8 165278 14 AC027380 Mus muscu
c 119 20.2 80.8 165316 14 AC119115 Rattus no
c 120 20.2 80.8 167168 15 AP005903 Oryza sat
c 121 20.2 80.8 167526 14 AC121622 Rattus no
c 122 20.2 80.8 167710 14 AC109891 Rattus no
c 123 20.2 80.8 168843 14 AC091711 Rattus no
c 124 20.2 80.8 169294 14 AC112587 Rattus no
c 125 20.2 80.8 169701 14 AC128113 Rattus no
c 126 20.2 80.8 170511 14 AC116263 Rattus no
c 127 20.2 80.8 171292 14 AC118493 Rattus no
c 128 20.2 80.8 171391 9 AC159372 Mus muscu
c 129 20.2 80.8 171817 14 AC027681 Mus muscu
c 130 20.2 80.8 172432 14 AC121221 Rattus no
c 131 20.2 80.8 172786 9 AC069469 Mus muscu
c 132 20.2 80.8 172872 9 AC140456 Mus muscu
c 133 20.2 80.8 173571 14 AC127725 Rattus no
c 134 20.2 80.8 173877 14 AC136010 Rattus no
c 135 20.2 80.8 175586 9 AC117807 Rattus no
c 136 20.2 80.8 175702 14 AC128541 Rattus no
c 137 20.2 80.8 175876 14 AC123336 Rattus no
c 138 20.2 80.8 176602 14 AC095858 Rattus no
c 139 20.2 80.8 177060 14 CR318634 Mus muscu
c 140 20.2 80.8 177227 14 AC117009 Rattus no
c 141 20.2 80.8 177653 9 AC159102 Mus muscu
c 142 20.2 80.8 177885 14 AC141947 Rattus no
c 143 20.2 80.8 178350 9 AC159621 Mus muscu
c 144 20.2 80.8 178671 9 AL805937 Mouse DNA
c 145 20.2 80.8 178815 14 AC127480 Rattus no
c 146 20.2 80.8 179019 9 AC103598 Mus muscu
c 147 20.2 80.8 179341 14 AC128908 Rattus no
c 148 20.2 80.8 179454 14 AC123737 Mus muscu
c 149 20.2 80.8 179481 14 AC034100 Mus muscu
c 150 20.2 80.8 180007 14 AC132933 Mus muscu
```

## ALIGNMENTS

```
RESULT 1
AC134563 AC134563 163878 bp DNA linear ROD 19-SEP-2003
LOCUS Mus musculus chromosome 19 clone RP24-179M18, complete sequence.
DEFINITION AC134563
ACCESSION AC134563
VERSION AC134563.5 GI:34850571
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 163878)  
Wilson,R.K.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 163878)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 163878)  
Wilson,R.K.  
Direct Submission  
Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 163878)  
Wilson,R.K.  
Direct Submission  
Submitted (19-SEP-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Sep 19, 2003 this sequence version replaced gi.34494974.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@wustl.edu  
----- Project Information -----  
Center project name: M\_BB0179M18  
-----

FEATURES  
source

Location/Qualifiers  
1..163878  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="19"  
/clone="RP24-179M18"

## ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 163878;  
Best Local Similarity 96.0%; Pred. No. 1.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAGAGCGAGACTGCTCTTC 25  
|||||  
Db 47914 AAAAAAGAGCGAGACTGCTCTTC 47938

## RESULT 2

AC142098 193811 bp DNA linear ROD 01-JAN-2004  
AC142098/c Mus musculus BAC clone RP24-458F14 from chromosome 19, complete  
sequence.  
DEFINITION AC142098  
ACCESSION AC142098  
VERSION AC142098.2 GI:38678651  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 193811)  
Veizer,J., Haakenson,W. and Bielicki,L.  
The sequence of Mus musculus BAC clone RP24-458F14  
Unpublished (2001)  
2 (bases 1 to 193811)  
Wilson,R.K.  
Sequencing of Mus musculus  
Unpublished (2001)  
3 (bases 1 to 193811)  
McPherson,J.D. and Waterston,R.H.

```
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
```

**TITLE**  
JOURNAL

**REFERENCE**  
AUTHORS

**TITLE**  
JOURNAL

**REFERENCE**  
AUTHORS

**TITLE**  
JOURNAL

**COMMENT**

Direct Submission  
Submitted (21-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 193811)  
Wilson.R.K.

Direct Submission  
Submitted (04-DEC-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 193811)  
Wilson.R.

Direct Submission  
Submitted (01-JAN-2004) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Dec 4, 2003 this sequence version replaced gi:29135708.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Summary Statistics  
Center project name: M\_BB0458F14  
-----

**NOTICE:** This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and  
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen  
and/or brain genomic DNA. The clone and detailed information can be  
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is  
overlapped by AC134563 and AC127271.

#### FEATURES

source  
1. .193811  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="19"  
/map="19"  
/clone="RP24-458F14"  
/clone\_lib="RPCI-24"  
26944\_-27849  
/note="CpG island (%GC=70.8, o/e=0.74, #CpGs=87)"  
62866..62738  
misc\_feature  
/product="trRNA-Ser"  
trRNA  
/note="Likely pseudogene (HMM Sc=40.10 / Sec struct  
Sc=12.19)"  
66009..66081  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=36.29 / Sec struct  
Sc=14.73)"  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=33.19 / Sec struct

Sc=-10.52)"  
complement(71547..71617)  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=30.48 / Sec struct  
Sc=-8.47)"  
complement(72364..72436)  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=39.40 / Sec struct  
Sc=-17.01)"  
75006..75077  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=30.13 / Sec struct  
Sc=-3.70)"  
complement(76074..76145)  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=35.88 / Sec struct  
Sc=-12.12)"  
misc\_feature  
88499..89380  
/note="CpG island (%GC=62.5, o/e=0.78, #CpGs=76)"  
complement(139660..139731)  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=33.09 / Sec struct  
Sc=-11.42)"  
complement(173218..173284)  
trRNA  
/product="trRNA-Lys"  
complement(175362..175434)  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=36.22 / Sec struct  
Sc=-14.60)"  
185833..185903  
trRNA  
/product="trRNA-Ser"  
/note="Likely pseudogene (HMM Sc=45.77 / Sec struct  
Sc=-12.81)"  
ORIGIN  
Query Match 93.6%; Score 23.4; DB 9; Length 193811;  
Best Local Similarity 96.0%; Pred. NO. 1.8;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGCTCTTCC 25  
Db 18517 AAAAAAGCGCAGACTGCTCTTCC 18493  
RESULT 3  
AC166153  
LOCUS  
DEFINITION  
AC166153 Mus musculus chromosome 7 clone w11-156D7 map 7, WORKING DRAFT  
SEQUENCES, 6 unordered pieces.  
AC166153.1 GI:71143326 linear HTG 26-JUL-2005  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
VERSION  
KEYWORDS  
SOURCE  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 28365)  
REFERENCE  
Birren, B., Nusbaum, C. and Lander, E.  
AUTHORS  
Mus musculus chromosome 7, clone w11-156D7  
TITLE  
Unpublished  
JOURNAL  
REFERENCE  
2 (bases 1 to 28365)  
AUTHORS  
Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V.,  
Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J.,  
Chopel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,  
Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D.,  
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,



McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,D., Raymond,C.,  
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Stubbs,M., Talanas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
 Zimmer,A. and Zody,M.

# TITLE JOURNAL

Submitted (26-JUL-2005) Broad Institute of MIT and Harvard, 320  
 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@broad.mit.edu

----- Project Information

Center project name: L33593

Center clone name: 156 D 7

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Assembly: Dye-terminator Big Dye; 100% of reads

Consensus quality: Phrap; version 0.960731

Consensus quality: 26029 bases at least Q40

Consensus quality: 26780 bases at least Q30

Consensus quality: 27410 bases at least Q20

Insert size: 40000; agarose-fp

Insert size: 27865; sum-of-contigs

Quality coverage: 9.4 in Q20 bases; agarose-fp

Quality coverage: 13.5 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1  
 \* 10487: contig of 10487 bp in length  
 \* 10488 10587: gap of unknown length  
 \* 10588 11249: contig of 662 bp in length  
 \* 11250 11349: gap of unknown length  
 \* 11350 12854: contig of 1505 bp in length  
 \* 12855 12954: gap of unknown length  
 \* 12955 14042: contig of 1088 bp in length  
 \* 14043 14142: gap of unknown length  
 \* 14143 15428: contig of 1286 bp in length  
 \* 15429 15528: gap of unknown length  
 \* 15529 28365: contig of 12837 bp in length.

## FEATURES

### source

1. .28365  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="7"  
 /map="7"  
 /clone="wi1-15607"  
 /clone\_lib="WIBR1 Mouse Fosmid Library"

misc\_feature  
 1. .10487  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"

gap  
 10488..10587  
 /estimated\_length=unknown

misc\_feature  
 10588..11249  
 /note="assembly\_fragment"

gap  
 11250..11349

misc\_feature  
 11350..12854  
 /note="assembly\_fragment"  
 gap  
 12855..12954  
 /estimated\_length=unknown  
 misc\_feature  
 12955..14042  
 /note="assembly\_fragment"  
 gap  
 14043..14142  
 /estimated\_length=unknown  
 misc\_feature  
 14143..15428  
 /note="assembly\_fragment"  
 gap  
 15429..15528  
 /estimated\_length=unknown  
 misc\_feature  
 15529..28365  
 /note="assembly\_fragment"

### ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 28365;  
 Best Local Similarity 92.0%; Pred. No. 11;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTGGTCTTCC 25

|||||||

Db 12298 AAAAAGAGCACTGACTGGTCTTCC 12322

### RESULT 4

CR788287

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

/estimated length=unknown  
 11350..12854  
 /note="assembly\_fragment"  
 12855..12954  
 /estimated length=unknown  
 12955..14042  
 /note="assembly\_fragment"  
 14043..14142  
 /estimated length=unknown  
 14143..15428  
 /note="assembly\_fragment"  
 15429..15528  
 /estimated length=unknown  
 15529..28365  
 /note="assembly\_fragment"

Mouse DNA sequence from clone RP23-362B9 on chromosome 2, complete  
 sequence.  
 CR788287  
 CR788287.7 GI:59796591  
 HTG.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 98311)  
 Direct Submission  
 Submitted (12-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk  
 On Feb 14, 2005 this sequence version replaced gi:58331666.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information

on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-362B9 is  
 from the RPCI-23 Mouse BAC library  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACE3.6.

## FEATURES

Location/Qualifiers  
 1..98311  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-362B9"  
 /clone\_lib="RPCI-23"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 98311;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTCTCC 25

Db 39405 AAAAAAGCAGCTGCTGGCTCTCC 39429

## RESULT 5

BX546458.1/c

## WPCOMMENT

Sequence split into 4 fragments LOCUS BX546458 Accession BX546458

Fragment Name	Begin	End
BX546458.0	1	110000
BX546458.1	100001	210000
BX546458.2	200001	310000
BX546458.3	300001	370067

Continuation (2 of 4) of BX546458 from base 100001 (BX546458 Mus musculus chromosome 2)

Query Match 87.2%; Score 21.8; DB 14; Length 110000;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTCTCC 25

Db 19233 AAAAAAGCAGCTGCTGGCTCTCC 19209

## RESULT 6

BX546462.0

## WPCOMMENT

Sequence split into 5 fragments LOCUS BX546462 Accession BX546462

Fragment Name	Begin	End
BX546462.0	1	110000
BX546462.1	100001	210000
BX546462.2	200001	310000
BX546462.3	300001	410000
BX546462.4	400001	447196

LOCUS BX546462 447196 bp DNA linear HTG 24-SEP-2003  
 DEFINITION Mus musculus chromosome 2 clone RP23-325022, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 38 unordered pieces.

## ACCESSION

BX546462

## VERSION

BX546462.3 GI:35209402

## KEYWORDS

HTG; HTGS PHASE1.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 447196)

Sims, S.

Direct Submission

Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 24, 2003 this sequence version replaced gi:32563023.

## COMMENT

Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

----- Project Information

Center project name: BM325022

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 434558 bases at least Q40

Consensus quality: 438112 bases at least Q30

Consensus quality: 440594 bases at least Q20

Insert size: 443496; sum-of-contigs

Insert size: 189301; 4.9% error; agarose-fp

Quality coverage: 5.28x in Q20 bases; sum-of-contigs Quality

coverage: 14.38x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1	14162:	contig of 14162 bp in length
*	14163:	gap of 100 bp
*	14263:	contig of 5397 bp in length
*	19659:	gap of 100 bp
*	19660:	contig of 2022 bp in length
*	21781:	gap of 100 bp
*	21881:	contig of 14529 bp in length
*	21882:	gap of 100 bp
*	36511:	contig of 5166 bp in length
*	36511:	gap of 100 bp
*	41677:	contig of 12447 bp in length
*	41677:	gap of 100 bp
*	41777:	contig of 100 bp
*	54224:	contig of 5350 bp in length
*	54324:	gap of 100 bp
*	59674:	contig of 7398 bp in length
*	59774:	gap of 100 bp
*	67172:	contig of 2187 bp in length
*	67172:	gap of 100 bp
*	69458:	contig of 19601 bp in length
*	69559:	gap of 100 bp
*	89160:	contig of 33372 bp in length
*	89260:	gap of 100 bp
*	122632:	contig of 11241 bp in length
*	122732:	gap of 100 bp
*	133973:	contig of 9393 bp in length
*	134073:	gap of 100 bp
*	143466:	contig of 44052 bp in length
*	143466:	gap of 100 bp
*	143566:	contig of 80458 bp in length
*	187618:	gap of 100 bp
*	187718:	contig of 40248 bp in length
*	268176:	gap of 100 bp
*	268276:	contig of 11639 bp in length
*	308524:	gap of 100 bp
*	308524:	contig of 3758 bp in length
*	320363:	gap of 100 bp
*	320363:	contig of 3641 bp in length
*	324121:	gap of 100 bp
*	324221:	contig of 2499 bp in length
*	327862:	gap of 100 bp
*	327862:	contig of 100 bp
*	330461:	gap of 100 bp
*	330560:	contig of 3453 bp in length
*	334013:	gap of 100 bp
*	334013:	contig of 4314 bp in length
*	334114:	gap of 100 bp
*	338427:	contig of 4314 bp in length

```
* 338428 338527: gap of 100 bp
* 338528 341755: contig of 3228 bp in length
* 341756 341855: gap of 100 bp
* 341856 346221: contig of 4366 bp in length
* 346222 346321: gap of 100 bp
* 346322 348507: contig of 2186 bp in length
* 348508 348607: gap of 100 bp
* 348608 378034: contig of 29427 bp in length
* 378035 378134: gap of 100 bp
* 378135 406759: contig of 28625 bp in length
* 406760 406859: gap of 100 bp
* 406860 411480: contig of 4621 bp in length
* 411481 411580: gap of 100 bp
* 411581 414369: contig of 2789 bp in length
* 414370 414469: gap of 100 bp
* 414470 418926: contig of 4457 bp in length
* 418927 419026: gap of 100 bp
* 419027 424085: contig of 5059 bp in length
* 424086 424185: gap of 100 bp
* 424186 426681: contig of 2496 bp in length
* 426682 426781: gap of 100 bp
* 426782 429890: contig of 3109 bp in length
* 429891 432321: contig of 2331 bp in length
* 432322 432421: gap of 100 bp
* 432422 435182: contig of 2761 bp in length
* 435183 435282: gap of 100 bp
* 435283 437359: contig of 2077 bp in length
* 437360 437459: gap of 100 bp
* 437460 444723: contig of 7264 bp in length
* 444724 444823: gap of 100 bp
* 444824 447196: contig of 2373 bp in length.
```

## FEATURES

## Source

```
1. .447196
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosomes="2"
  /clone_lib="RPCI-23"
  /clone="RP23-325O22"
  1. .14162
    /note="assembly fragment:03333
    fragment_chain:1"
  14263..19659
    /note="assembly fragment:04136
    fragment_chain:1"
  19760..21781
    /note="assembly fragment:01430
    fragment_chain:1"
  21882..36410
    /note="assembly fragment:04198
    fragment_chain:1"
  36511..41676
    /note="assembly fragment:02008
    fragment_chain:1"
  41777..54223
    /note="assembly fragment:04185
    fragment_chain:1"
  54324..59673
    /note="assembly fragment:01187
    fragment_chain:1"
  59774..67171
    /note="assembly fragment:02108
    fragment_chain:1"
  67272..69458
    /note="assembly fragment:04263
    fragment_chain:1"
  69559..89159
    /note="assembly fragment:00095
    fragment_chain:2"
  89260..122631
    /note="assembly fragment:03458
    fragment_chain:2"
```

```
misc_feature 122732..133972
              /note="assembly fragment:02236
              fragment_chain:2"
misc_feature 134073..143465
              /note="assembly fragment:03351
              fragment_chain:3"
misc_feature 143566..187617
              /note="assembly fragment:02065
              fragment_chain:3"
misc_feature 187718..288175
              /note="assembly fragment:03774
              fragment_chain:3"
misc_feature 268276..308523
              /note="assembly fragment:03651
              fragment_chain:4"
misc_feature 308624..320262
              /note="assembly fragment:04052
              fragment_chain:4"
misc_feature 320363..324120
              /note="assembly fragment:00046"
              fragment_chain:4"
misc_feature 324221..327861
              /note="assembly fragment:00226"
              fragment_chain:4"
misc_feature 327962..330460
              /note="assembly fragment:00242"
              fragment_chain:4"
misc_feature 330561..334013
              /note="assembly fragment:00624"
              fragment_chain:4"
misc_feature 334114..338427
              /note="assembly fragment:00667"
              fragment_chain:4"
```

Query Match 87.2%; Score 21.8; DB 14; Length 110000;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGCTGCTTCC 25

Db 88584 AAAAAAGAGACTGACTGCTTCC 88608

RESULT 7  
CR753898\_0/c  
WPCOMMENT

Sequence split into 5 fragments LOCUS CR753898 Accession CR753898

Fragment Name	Begin	End
CR753898_0	1	110000
CR753898_1	100001	210000
CR753898_2	200001	310000
CR753898_3	300001	410000
CR753898_4	400001	427662

LOCUS CR753898 427662 bp DNA linear HTG 10-OCT-2004  
DEFINITION Mus musculus chromosome 2 clone RP23-100D9, 2 unordered pieces.  
ACCESSION CR753898  
VERSION CR753898.5 GI:54021828  
KEYWORDS HTG; HTGS PHASE1; HTGS CANCELLED.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 427662)

AUTHORS Wood,J.

TITLE Direct Submission

JOURNAL Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Oct 10, 2004 this sequence version replaced gi:53850270.  
Sequence from the Mouse Genome Sequencing Consortium whole genome  
shotgun may have been used to confirm this sequence. Sequence data  
from the whole genome shotgun alone has only been used where it has  
a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: bM100D9  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: dye-terminator; 100% of reads  
 Consensus quality: 201139 bases at least Q40  
 Consensus quality: 201243 bases at least Q30  
 Consensus quality: 204170 bases at least Q20  
 Insert size: 427562; sum-of-contigs  
 Insert size: 200670; 4.2% error; agarose-fp  
 Quality coverage: 3.82x in Q20 bases; sum-of-contigs Quality  
 coverage: 8.39x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 2131: contig of 2131 bp in length  
 \* 2132 2231: gap of 100 bp  
 \* 2232 427662: contig of 425431 bp in length.  
 -----  
 FEATURES  
 source  
 Location/Qualifiers  
 1..427662  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-100D9"  
 /clone\_lib="RPCI-23"  
 1..2131  
 /note="assembly\_fragment:00090"  
 2232..427662  
 /note="assembly\_fragment:02493"  
 clone\_end:T7  
 vector\_side:right  
 misc\_feature  
 misc\_feature  
 2232..427662  
 /note="assembly\_fragment:02493"  
 clone\_end:T7  
 vector\_side:right  
 ORIGIN  
 Query Match 87.2%; Score 21.8; DB 14; Length 110000;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAAAAAGAGCGACTGGTCTTCC 25  
 Db 53218 AAAAAAGAGCGACTGGTCTTCC 53194  
 RESULT 8  
 CR753898\_2  
 WPCOMMENT  
 Sequence split into 5 fragments LOCUS CR753898 Accession CR753898  
 Fragment Name Begin End  
 CR753898\_0 1 110000  
 CR753898\_1 100001 210000  
 CR753898\_2 200001 310000  
 CR753898\_3 300001 410000  
 CR753898\_4 400001 427662  
 Continuation (3 of 5) of CR753898 Mus musculus chromosome 2  
 Query Match 87.2%; Score 21.8; DB 14; Length 110000;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAAAAAGAGCGACTGGTCTTCC 25  
 Db 19431 AAAAAAGAGCGACTGGTCTTCC 19455  
 RESULT 9  
 BX284639  
 LOCUS  
 DEFINITION  
 Mouse DNA sequence from clone RP24-87L14 on chromosome 2, complete  
 sequence.  
 ACCESSION  
 BX284639  
 VERSION  
 BX284639.15 GI:37805733  
 KEYWORDS  
 HTG.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 115602)  
 Howden,P.  
 DIRECT SUBMISSION  
 Submitted (21-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk  
 Clonerequests@sanger.ac.uk  
 On Oct 21, 2003 this sequence version replaced gi:37665507.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun map have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/ RP24-87L14 is  
 from a Male (C57BL/6J) mouse BAC Library VECTOR: PTARBAC1.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..115602  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP24-87L14"  
 /clone\_lib="RPCI-24"  
 ORIGIN  
 Query Match 87.2%; Score 21.8; DB 9; Length 115602;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAAAAAGAGCGACTGGTCTTCC 25  
 Db 15863 AAAAAAGAGCGACTGGTCTTCC 15887  
 RESULT 10  
 CR974476/c  
 LOCUS  
 DEFINITION  
 Mus musculus chromosome 2 clone RP24-469J5, WORKING DRAFT SEQUENCE,  
 2 unordered pieces.  
 ACCESSION  
 CR974476

CR974476.3 GI:71891486  
HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 136564)  
Almeida, J.  
Direct Submission  
Submitted (04-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegasanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Aug 5, 2005 this sequence version replaced gi:71833829.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegasanger.ac.uk  
----- Project Information  
Center project name: bN469J5  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 136421 bases at least Q40  
Consensus quality: 136424 bases at least Q30  
Consensus quality: 136439 bases at least Q20  
Insert size: 136464; sum-of-contigs  
Quality coverage: 12.82x in Q20 bases; agarose-fp  
Quality coverage: 11.95x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 44305: contig of 44305 bp in length  
\* 44306 44405: gap of 100 bp  
\* 44406 136564: contig of 92159 bp in length.

FEATURES

source  
1..136564  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone="RP24-459J5"  
/clone\_lib="RPC1-24"  
1..44305  
/note="assembly fragment:00898  
fragment chain:1  
clone end:SP6  
vector\_side:left"  
44406..136564  
/note="assembly fragment:00818  
fragment chain:1  
clone end:T7  
vector\_side:right"

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 136564;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAAAAAGACGACACTGCTCTTCC 25  
|||||  
DB 98063 AAAAAAGACGACACTGCTCTTCC 98039  
|||||

RESULT 11

AC118310  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

AC118310 143116 bp DNA linear HTG 15-NOV-2002  
Rattus norvegicus clone CH230-373K19, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC118310  
AC118310.5 GI:25009273  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 143116)

Murny D.Marie., Metzker M.Lee., Abramson S., Adams C., Alder J.,  
Allen C., Allen H., Alsbrooks S., Amin A., Angiano D.,  
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,  
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,  
Bigwalo K., Blair J., Blankenburg K., Blyth P., Brown M.,  
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,  
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,  
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,  
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,  
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,  
Deigado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,  
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval S., Eaves K.,  
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,  
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,  
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,  
Gebregeorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,  
Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,  
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,  
Hernandez R., Hines S., Hladun S.L., Hodgson A., Hogues M.,  
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,  
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jollivet A.,  
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,  
Kowis C., Kraft C., Lebow H., Levan J., Lewis L., Li Z., Liu J.,  
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,  
Lorensuewa L., Loulsegged H., Lozado R.J., Lu X., Ma J.,  
Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,  
Mangum B., Mapua P., Martin K., Martin R., Martinez E.,  
Mawhiney S., McLeod M.P., McNeill T.Z., Meenen E.,  
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,  
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Naif L.,  
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,  
Nwackeleme O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,  
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,  
Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,  
Puazo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,  
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,  
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,  
Sanders W., Savery G., Scherer S., Scott G., Shatsman S., Shen H.,  
Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Smajs D.,  
Snead A., Sodergren E., Song X.-Z., Sorelle R., Soza J.,  
Steinle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,  
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Uemani K.,  
Valas R., Vera V., Villanana D., Waldron L., Walker B., Wang J.,  
Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,  
Williams G., Willison R., Wleciyk R., Wooden H., Worley K.,  
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,  
Yu F., Zhang J., Zhou X., Zhou X., Zhao S., Dunn D., von  
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,  
Weinstock G. and Gibbs R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 143116)  
Worley K.C.  
Direct Submission  
Submitted (15-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 143116)  
Rat Genome Sequencing Consortium.  
Direct Submission

## JOURNAL

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Nov 15, 2002 this sequence version replaced gi:23195560. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information

Center project name: GVDU  
Center clone name: CH230-373K19

----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 136117 bases at least Q40  
Consensus quality: 137015 bases at least Q30  
Consensus quality: 137564 bases at least Q20  
Estimated insert size: 139703; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will be preserved.

\* 1 123378: contig of 123378 bp in length  
\* 123379 123478: gap of unknown length  
\* 123479 143116: contig of 19638 bp in length.

## FEATURES

source 1.143116  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-373K19"  
complement(362..1303)  
/note="clone boundary  
clone\_end:T7  
site:  
end sequence:B237824"

## misc\_feature

1841..2992  
/note="clone boundary  
clone\_end:Sp6  
site:  
end sequence:B237827"

## misc\_feature

123379..123478  
/estimated\_length=unknown

## gap

## ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 143116;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAAAAGAGCGGACTGGTCTTCC 25  
|||||  
DB 17089 AAAAAAGAGCGGACTGGTCTTCC 17113

RESULT 12  
AC166152/cLOCUS  
DEFINITIONACCESSION  
VERSIONKEYWORDS  
SOURCE

## ORGANISM

REFERENCE  
AUTHORSTITLE  
JOURNALREFERENCE  
AUTHORS

AC166152 143176 bp DNA linear HTG 26-JUL-2005  
Mus musculus chromosome 15 clone RP24-334D17 map 15, WORKING DRAFT  
SEQUENCE, 3 unordered pieces.  
AC166152  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.  
1 (bases 1 to 143176)  
Birren,B., Nuebaum,C. and Lander,B.

Mus musculus chromosome 15, clone RP24-334D17

Unpublished

2 (bases 1 to 143176)

Birren,B., Nuebaum,C., Lander,B., Abouelleil,A., Allen,N.,  
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,  
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,  
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,  
DeRellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,  
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,  
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,  
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,  
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,  
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R.,  
Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,  
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,  
Zimmer,A. and Zody,M.

Direct Submission

Submitted (26-JUL-2005) Broad Institute of MIT and Harvard, 320  
Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Broad Institute of MIT and Harvard

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@broad.mit.edu](mailto:sequence_submissions@broad.mit.edu)

----- Project Information

Center project name: L33591

Center clone name: 334\_D17

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 142838 bases at least Q40

Consensus quality: 142924 bases at least Q30

Consensus quality: 142955 bases at least Q20

Insert size: 200000; agarose-fp

Insert size: 142976; sum-of-contigs

Quality coverage: 11.4 in Q20 bases; agarose-fp

Quality coverage: 15.9 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will

\* be preserved.  
 \* 1 56560: contig of 56560 bp in length  
 \* 56561 56660: gap of unknown length  
 \* 56615 66715: contig of 9955 bp in length  
 \* 66616 66715: gap of unknown length  
 \* 66716 143176: contig of 76461 bp in length.

#### FEATURES

source  
 1..143176  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="15"  
 /map="15"  
 /clone="RP24-334D17"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 misc\_feature  
 1..56560  
 /note="assembly\_fragment  
 clone end:SP6  
 vector\_side:left"  
 56561..56660  
 /estimated\_length=unknown  
 56661..66615  
 /note="assembly\_fragment"  
 66616..66715  
 /estimated\_length=unknown  
 66716..143176  
 /note="assembly\_fragment  
 clone end:T7  
 vector\_side:right"

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 143176;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGCTGGTCTTCC 25  
 |||||  
 Db 51339 AAAAAAGAGCTGACTGGTCTTCC 51315

#### RESULT 13

CR3544442  
 LOCUS CR3544442 145110 bp DNA linear ROD 19-APR-2004  
 DEFINITION Mouse DNA sequence from clone RP24-562O19 on chromosome 2, complete  
 sequence.  
 ACCESSION CR354442  
 VERSION CR354442.5 GI:46408806  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 145110)  
 AUTHORS Palmer,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-APR-2004) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Apr 19, 2004 this sequence version replaced gi:46357668.  
 COMMENT Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

#### FEATURES

Location/Qualifiers  
 source  
 1..145110  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP24-562O19"  
 /clone\_lib="RPCI-24"

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 145110;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGCTGGTCTTCC 25  
 |||||

Db 16996 AAAAAAGAGCTGACTGGTCTTCC 17020

#### RESULT 14

CR3544442/c  
 LOCUS CR3544442 145110 bp DNA linear ROD 19-APR-2004  
 DEFINITION Mouse DNA sequence from clone RP24-562O19 on chromosome 2, complete  
 sequence.  
 ACCESSION CR354442  
 VERSION CR354442.5 GI:46408806  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 145110)  
 AUTHORS Palmer,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-APR-2004) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Apr 19, 2004 this sequence version replaced gi:46357668.  
 COMMENT Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all



regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP24-562019 is from a Male (C57Bl/6J) mouse BAC library VECTOR: pTARBAC1.

## FEATURES

source

Location/Qualifiers

```
1..145110
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP24-562019"
/clone_lib="RPCI-24"
```

## ORIGIN

```
Query Match      87.2%; Score 21.8; DB 9; Length 145110;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 AAAAAAGCGCAGCTGCTCTCC 25

|||||

Db 129932 AAAAAAGCAGCTGCTCTCC 129908

## RESULT 15

CT010482/c

LOCUS

DEFINITION

Mus musculus chromosome 17 clone RP23-77N9, WORKING DRAFT SEQUENCE, 3 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.  
1 (bases 1 to 145376)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)  
Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Aug 12, 2005 this sequence version replaced gi:71142041.

COMMENT

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)  
Project Information  
Center project name: bM77N9  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 144894 bases at least Q40  
Consensus quality: 145065 bases at least Q30  
Consensus quality: 145106 bases at least Q20  
Insert size: 145176; sum-of-contigs  
Insert size: 196719; 7.0% error; agarose-fp  
Quality coverage: 10.91x in Q20 bases; sum-of-contigs Quality coverage: 8.36x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 55513: contig of 55513 bp in length
* 55514 55613: gap of 100 bp
* 55614 142547: contig of 86934 bp in length
* 142548 142647: gap of 100 bp
* 142648 145376: contig of 2729 bp in length.
```

## FEATURES

source

Location/Qualifiers

```
1..145376
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-77N9"
/clone_lib="RPCI-23"
1..55513
/note="assembly_fragment:00038"
fragment chain:1
clone_end:SP6
vector_side:left"
55614..142547
/note="assembly_fragment:00845"
fragment chain:1"
142648..145376
/note="assembly_fragment:00030"
```

misc\_feature

misc\_feature

misc\_feature

## ORIGIN

```
Query Match      87.2%; Score 21.8; DB 14; Length 145376;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 AAAAAAGCGCAGCTGCTCTCC 25

|||||

Db 85674 AAAAAAGCAGCTGCTCTCC 85650

## RESULT 16

AL607123

LOCUS

DEFINITION

Mouse DNA sequence from clone RP23-391E6 on chromosome 4, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL607123 149369 bp DNA linear ROD 25-APR-2002  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.  
1 (bases 1 to 149369)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (25-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Apr 29, 2002 this sequence version replaced gi:20068562.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

COMMENT

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> RP23-391E6 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6 Location/Qualifiers

```

FEATURES
    source
        1..149369
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="4"
            /clone="RP23-391E6"
            /clone_lib="RPCI-23"

```

#### ORIGIN

```

Query Match      87.2%; Score 21.8; DB 9; Length 149369;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  AAAAAAGACGACACTGCTCTTCC 25
        |||||
Db      19283 AAAAAAGACACAGATTGCTCTTCC 19307

```

#### RESULT 17

```

CR955032/c
LOCUS      151693 bp DNA linear HTG 25-MAY-2005
DEFINITION Homo sapiens clone CHORI-516-BCA274P22, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.

```

```

ACCESSION  CR955032.2 GI:66711321
VERSION     HTG; HTGS PHASE1.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens

```

#### ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```

Sims, S.
1 (bases 1 to 151693)

```

#### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

#### COMMENT

```

Direct Submission
Submitted (24-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 25, 2005 this sequence version replaced gi:66392756.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bCA274P22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 151132 bases at least Q40
Consensus quality: 151298 bases at least Q30
Consensus quality: 151407 bases at least Q20
Insert size: 151493; sum-of-contigs
Insert size: 187370; 3.1% error; agarose-fp
Quality coverage: 11.97x in Q20 bases; sum-of-contigs Quality
coverage: 9.90x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

*      1  56088: contig of 56088 bp in length
*      *  56089  56188: gap of 100 bp
*      *  56189  115116: contig of 58928 bp in length
*      *  115117  151216: gap of 100 bp
*      *  115217  151693: contig of 36477 bp in length.
FEATURES
    Location/Qualifiers
        1..151693
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="CHORI-516-BCA274P22"
            /clone_lib="CHORI-516"

```

```

        1..56088
            /note="assembly_fragment:00606"
            fragment_chain:1
            clone_end:17
            vector_side:left"

```

```

        56189..115116
            /note="assembly_fragment:01538"
            fragment_chain:1
            115217..151693
            /note="assembly_fragment:00047"
            fragment_chain:1
            clone_end:SP6
            vector_side:right"

```

#### misc\_feature

```

        1..56088
            /note="assembly_fragment:00606"
            fragment_chain:1
            clone_end:17
            vector_side:left"

```

```

        56189..115116
            /note="assembly_fragment:01538"
            fragment_chain:1
            115217..151693
            /note="assembly_fragment:00047"
            fragment_chain:1
            clone_end:SP6
            vector_side:right"

```

```

        115217..151693
            /note="assembly_fragment:00047"
            fragment_chain:1
            clone_end:SP6
            vector_side:right"

```

#### misc\_feature

```

        115217..151693
            /note="assembly_fragment:00047"
            fragment_chain:1
            clone_end:SP6
            vector_side:right"

```

#### ORIGIN

```

Query Match      87.2%; Score 21.8; DB 14; Length 151693;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  AAAAAAGACGACACTGCTCTTCC 25
        |||||
Db      25889 AAAAAAGACACTGCTCTTCC 25865

```

#### RESULT 18

```

BX511235/c
LOCUS      169581 bp DNA linear ROD 24-OCT-2003
DEFINITION Mouse DNA sequence from clone RP23-232H13 on chromosome 2, complete
            sequence.

```

```

ACCESSION  BX511235
VERSION     BX511235.10 GI:37988139
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)

```

#### ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 169581)
Wood, J.
Direct Submission
Submitted (24-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 24, 2003 this sequence version replaced gi:37805699.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

```

#### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

#### COMMENT

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 151132 bases at least Q40
Consensus quality: 151298 bases at least Q30
Consensus quality: 151407 bases at least Q20
Insert size: 151493; sum-of-contigs
Insert size: 187370; 3.1% error; agarose-fp
Quality coverage: 11.97x in Q20 bases; sum-of-contigs Quality
coverage: 9.90x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-232H13 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6. Location/Qualifiers

```

FEATURES             Location/Qualifiers
     source            1..169581
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosomes="2"
                        /clone="RP23-232H13"
                        /clone_lib="RPCI-23"

```

## ORIGIN

```

Query Match          87.2%; Score 21.8; DB 9; Length 169581;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  AAAAAAGAGCGCAGACTGGTCTTCC 25
        |||||
Db       61942 AAAAAAGAGCAGTGTGCTTCC 61918

```

## RESULT 19

```

CR955031             174479 bp  DNA  linear  HTG 25-MAY-2005
LOCUS               Homo sapiens clone CHORI-516-BCA362L20, *** SEQUENCING IN PROGRESS
DEFINITION          ***, 5 unordered pieces.

```

```

ACCESSION          CR955031.2  GI:66711320
VERSION            CR955031.2  GI:66711320
KEYWORDS           HTG; HTGS_PHASE1.
SOURCE             Homo sapiens (human)
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                   Hominiidae; Homo.
REFERENCE          1 (bases 1 to 174479)
AUTHORS            Sims,S.

```

```

TITLE              Direct Submission
JOURNAL            Submitted (24-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
                   Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                   humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT            On May 25, 2005 this sequence version replaced gi:66392755.

```

## COMMENT

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BCA362L20

```

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 173057 bases at least Q40
Consensus quality: 173562 bases at least Q30
Consensus quality: 173854 bases at least Q20
Insert size: 174079; sum-of-contigs
Insert size: 207259; 2.8% error; agarose-fp
Quality coverage: 6.78x in Q20 bases; sum-of-contigs
coverage: 5.79x in Q20 bases; agarose-fp
-----

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 80118: contig of 80118 bp in length
* 80119 80218: gap of 100 bp
* 80219 147820: contig of 67602 bp in length
* 147821 147820: gap of 100 bp
* 147821 156970: contig of 9050 bp in length
* 147921 156970: gap of 100 bp
* 156971 157070: gap of 100 bp
* 157071 165081: contig of 8011 bp in length
* 157071 165181: gap of 100 bp
* 165082 174479: contig of 9298 bp in length.

```

## FEATURES

## source

```

1..174479
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="CHORI-516-BCA362L20"
/clone_lib="CHORI-516"

```

## misc\_feature

```

1..80118
/note="assembly_fragment:00975"
fragment_chain:1"

```

## misc\_feature

```

80219..147820
/note="assembly_fragment:00283"
fragment_chain:1"

```

## misc\_feature

```

147921..156970
/note="assembly_fragment:00186"
fragment_chain:1"

```

## misc\_feature

```

157071..165081
/note="assembly_fragment:00034"
fragment_chain:2"

```

## misc\_feature

```

165182..174479
/note="assembly_fragment:00104"
fragment_chain:2"

```

```

clone_end:T7
vector_side:right"

```

## ORIGIN

```

Query Match          87.2%; Score 21.8; DB 14; Length 174479;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  AAAAAAGAGCGCAGACTGGTCTTCC 25
        |||||
Db       119987 AAAAAAGAGCAGTGTGCTTCC 120011

```

## RESULT 20

## CR955031/c

## LOCUS

## DEFINITION

```

Homo sapiens clone CHORI-516-BCA362L20, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.

```

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: BCA362L20

----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 173057 bases at least Q40  
 Consensus quality: 173562 bases at least Q30  
 Consensus quality: 173854 bases at least Q20  
 Insert size: 174079; sum-of-contigs  
 Insert size: 207259; 2.8% error; agarose-fp  
 Quality coverage: 6.78x in Q20 bases; sum-of-contigs Quality  
 coverage: 5.79x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 80118: contig of 80118 bp in length  
 \* 80119 80218: gap of 100 bp  
 \* 80219 147820: contig of 67602 bp in length  
 \* 147821 147920: gap of 100 bp  
 \* 147921 156970: contig of 9050 bp in length  
 \* 156971 157070: gap of 100 bp  
 \* 157071 165081: contig of 8011 bp in length  
 \* 165082 165181: gap of 100 bp  
 \* 165182 174479: contig of 9298 bp in length.

## FEATURES

source

1..174479  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="CHORI-516-BCA362L20"  
 /clone\_lib="CHORI-516"

misc\_feature

1..80118  
 /note="assembly fragment:00975  
 fragment\_chain:1"

misc\_feature

80219..147820  
 /note="assembly fragment:00283  
 fragment\_chain:1"

misc\_feature

147921..156970  
 /note="assembly fragment:00186  
 fragment\_chain:1"

misc\_feature

157071..165081  
 /note="assembly fragment:00034  
 fragment\_chain:2"

misc\_feature

165182..174479  
 /note="assembly fragment:00104  
 fragment\_chain:2  
 clone\_end:T7  
 vector\_side:right"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 174479;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGGTCTTCC 25  
 |||||

Db 165208 AAAAAAGAGACTGACTGGTCTTCC 165184  
 |||||

RESULT 21

BX005149

LOCUS

DEFINITION

BX005149 179228 bp DNA linear ROD 18-JUL-2003  
 Mouse DNA sequence from clone RP24-574J13 on chromosome 2, complete

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

sequence.  
 BX005149  
 HTG.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 179228)  
 Tracey,A.  
 Direct Submission  
 Submitted (18-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequests@sanger.ac.uk](mailto:clonerequests@sanger.ac.uk)  
 On Jul 18, 2003 this sequence version replaced gi:32487195.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP24-574J13 is  
 from a Male (C57BL/6J) mouse BAC Library VECTOR: pIARBAC1.

## FEATURES

source

Location/Qualifiers  
 1..179228  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP24-574J13"  
 /clone\_lib="RPCI-24"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 179228;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGGTCTTCC 25  
 |||||

Db 6449 AAAAAAGAGACTGACTGGTCTTCC 6473  
 |||||

RESULT 22

BX005149/c

LOCUS

DEFINITION

BX005149 179228 bp DNA linear ROD 18-JUL-2003  
 Mouse DNA sequence from clone RP24-574J13 on chromosome 2, complete  
 sequence.

ACCESSION

BX005149

VERSION

KEYWORDS

BX005149.16 GI:32968213  
 HTG.

```

SOURCE          Mus musculus (house mouse)
ORGANISM
REFERENCE       Mus musculus
AUTHORS         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL         Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 179228)
Direct Submission
Submitted (18-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 18, 2003 this sequence version replaced gi:32487195.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep/
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
1..179228
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP24-574J13"
/clone_lib="RPCI-24"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 179228;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
|||||
Db 119388 AAAAAAGAGCGCAGCTGACTGGTCTTCC 119364

RESULT 23
AL845456/c
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-90C6 on chromosome 2, complete
sequence.
ACCESSION
AL845456
VERSION
AL845456.10 GI:28445893
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

SOURCE          Mus musculus (house mouse)
ORGANISM
REFERENCE       Mus musculus
AUTHORS         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL         Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 180947)
Direct Submission
Submitted (25-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 20, 2003 this sequence version replaced gi:27817366.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep/
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
1..180947
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-90C6"
/clone_lib="RPCI-23"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 180947;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
|||||
Db 142630 AAAAAAGAGCGCTGACTGGTCTTCC 142606

RESULT 24
BX682537
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-376N23 on chromosome 2, complete
sequence.
ACCESSION
BX682537
VERSION
BX682537.3 GI:38524372
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

```

REFERENCE 1 (bases 1 to 186050)  
AUTHORS Sycamore,N.  
TITLE Direct Submission  
JOURNAL Submitted (25-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Nov 25, 2003 this sequence version replaced gi:38153729.  
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun map have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-376N23 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6.  
-----  
FEATURES  
source  
1. 186050  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="2"  
/clone="RP23-376N23"  
/clone\_lib="RPCI-23"  
-----  
ORIGIN  
Query Match 87.2%; Score 21.8; DB 9; Length 186050;  
Best Local Similarity 92.0%; Pred. NO. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
-----  
QY 1 AAAAAAGCGAGCTGCTCTCC 25  
Db 131777 AAAAAAGAGCACTGACTGCTCTCC 131801  
-----  
RESULT 25  
AC118842  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-245G5, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
AC118842  
AC118842 190928 bp DNA linear HTG 19-NOV-2002  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
1 (bases 1 to 190928)

## AUTHORS

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Delgado,M., Denis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuheva,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munitasa,M., Murphy,M., Nait,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfanckoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shivartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 190928)  
Worley,K.C.  
Direct Submission  
Submitted (21-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 190928)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23908255.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole



Genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVUO
Center clone name: CH230-245G5
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 182925 bases at least Q40
Consensus quality: 183784 bases at least Q30
Consensus quality: 184265 bases at least Q20
Estimated insert size: 188977; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 190928: contig of 190928 bp in length.
FEATURES
source
1..190928
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-245G5"
misc_feature
1..1950
/notes="wgs contig"
misc_feature
2377..3683
/notes="wgs contig"
misc_feature
182872..183783
/notes="clone boundary"
clone_end:Sp6
site:
end sequence:B2140272"
184283..186093
/notes="wgs end extension"
clone_end:Sp6"
misc_feature
188925..190928
/notes="wgs_end_extension"
clone_end:Sp6"

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 190928;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGGTCTTCC 25
|||||
Db 138712 AAAAAAGAGCGACACTGGTCTTCC 138736
|||||

RESULT 26
CR318639 193326 bp DNA linear ROD 18-MAY-2004
LOCUS Mouse DNA sequence from clone RP23-317N11 on chromosome 2, complete
DEFINITION sequence.
ACCESSION CR318639
VERSION CR318639.6 GI:47496683
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

# REFERENCE

## AUTHORS

### JOURNAL

# COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 193326)

Almeida,J.

Direct Submission

Submitted (18-MAY-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On May 18, 2004 this sequence version replaced gi:46934930.

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Swt:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-317N11 is from the RPCI-23 Mouse BAC library constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6.

# FEATURES

## source

```

1..193326
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="2"
/clone="RP23-317N11"
/clone_lib="RPCI-23"

```

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 193326;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGGTCTTCC 25

Db 178491 AAAAAAGAGCGACACTGGTCTTCC 178515

## RESULT 27

EX465218 196469 bp DNA linear HTG 20-MAY-2003

LOCUS Mus musculus chromosome 2 clone RP23-419P6, 3 unordered pieces.

DEFINITION BX465218

ACCESSION BX465218

VERSION BX465218.3 GI:30962340

KEYWORDS HTG; HTGS PHASE1; HTGS\_CANCELLED.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;



```
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Sciurognathi; Murioidea; Muridae; Murinae; Mus.
1 (bases 1 to 196469)
Sycamore,N.
Direct Submission
Submitted (18-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 20, 2003 this sequence version replaced gi:30842655.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm419P6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 191269 bases at least Q40
Consensus quality: 191438 bases at least Q30
Consensus quality: 191523 bases at least Q20
Insert size: 196269; sum-of-contigs
Insert size: 203083; 4.2% error; agarose-fp
Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
coverage: 6.83x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 60380: contig of 60380 bp in length
* 60381 60480: gap of 100 bp
* 60481 72206: contig of 11726 bp in length
* 72207 72306: gap of 100 bp
* 72307 196469: contig of 124163 bp in length.
-----
Location/Qualifiers
1. .196469
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="2"
/clone_lib="RPCI-23"
/clone_id="RP23-419P6"
misc_feature
1. .60380
/note="assembly_fragment:02020
clone_end:SP6
vector_side:left"
misc_feature
60481..72206
/note="assembly_fragment:02063
fragment_chain:1"
misc_feature
72307..196469
/note="assembly_fragment:02021
fragment_chain:1"
ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 196469;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGGAGACTGGTCTTCC 25
|||||
Db 142575 AAAAAAGAGACTGACTGGTCTTCC 142599
|||||

FEATURES
source
1. .196469
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="2"
/clone_lib="RPCI-23"
/clone_id="RP23-419P6"
misc_feature
1. .60380
/note="assembly_fragment:02020
clone_end:SP6
vector_side:left"
misc_feature
60481..72206
/note="assembly_fragment:02063
fragment_chain:1"
misc_feature
72307..196469
/note="assembly_fragment:02021
fragment_chain:1"
ORIGIN

CR956625/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CR956625
Mus musculus chromosome 2 clone RP23-100C10, *** SEQUENCING IN
PROGRESS ***, 13 unordered pieces.
CR956625
GI:67509302
HTG; HTGS_PHASE1.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 198262)
Sims,S.
Direct Submission
Submitted (08-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jun 10, 2005 this sequence version replaced gi:66947690.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
----- Project Information
Center project name: bm100C10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 193007 bases at least Q40
Consensus quality: 194081 bases at least Q30
Consensus quality: 195094 bases at least Q20
Insert size: 197062; sum-of-contigs
Insert size: 195467; 5.4% error; agarose-fp
Quality coverage: 4.75x in Q20 bases; sum-of-contigs Quality
coverage: 4.82x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 55953: contig of 55953 bp in length
* 55954 56053: gap of 100 bp
* 56054 63707: contig of 7654 bp in length
* 63708 63807: gap of 100 bp
* 63808 66578: contig of 2771 bp in length
* 66579 66678: gap of 100 bp
* 66679 104450: contig of 37772 bp in length
* 104451 104550: gap of 100 bp
* 104551 116967: contig of 12417 bp in length
* 116968 117067: gap of 100 bp
* 117068 126883: contig of 9816 bp in length
* 126884 126983: gap of 100 bp
* 126984 143917: contig of 16934 bp in length
* 143918 144017: gap of 100 bp
* 144018 147138: contig of 3121 bp in length
* 147139 147238: gap of 100 bp
* 147239 165105: contig of 17867 bp in length
* 165106 165205: gap of 100 bp
* 165206 180021: contig of 14816 bp in length
* 180022 180121: gap of 100 bp
* 180122 187202: contig of 7081 bp in length
* 187203 187302: gap of 100 bp
* 187303 190040: contig of 2738 bp in length
* 190041 190140: gap of 100 bp
* 190141 198262: contig of 8122 bp in length.
-----
Location/Qualifiers
1. .198262
/organism="Mus musculus"
```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

```

/misc_feature  /mol_type="genomic DNA"
               /db_xref="taxon:10090"
               /chromosome="2"
               /clone="RP23-100C10"
               /clone_lib="RPCI-23"
               1. 55953
               /notes="assembly fragment: 00978"
               fragment chain: 1
               clone_end: T7
               vector_side: left"
               56054. .63707
               /notes="assembly fragment: 00094"
               fragment chain: 1"
               63808. .66578
               /notes="assembly fragment: 00032"
               fragment chain: 1"
               66679. .104450
               /notes="assembly fragment: 00699"
               fragment chain: 1"
               104551. .116967
               /notes="assembly fragment: 00259"
               fragment chain: 1"
               117068. .126883
               /notes="assembly fragment: 00195"
               fragment chain: 1"
               126984. .143917
               /notes="assembly fragment: 00576"
               fragment chain: 1"
               144018. .147138
               /notes="assembly fragment: 00018"
               fragment chain: 1"
               147239. .165105
               /notes="assembly fragment: 00463"
               fragment chain: 1"
               165206. .180021
               /notes="assembly fragment: 00357"
               fragment chain: 1"
               180122. .187202
               /notes="assembly fragment: 00142"
               fragment chain: 1"
               187303. .190040
               /notes="assembly fragment: 00009"
               fragment chain: 1"
               clone_end: SP6
               vector_side: right"
               190141. .198262
               /notes="assembly fragment: 00047"

```

ORIGIN

```

Query Match      87.2%; Score 21.8; DB 14; Length 198262;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1  AAAAAAGCGCAGCTGCTTCC 25
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db   113064 AAAAAAGCGCAGCTGCTTCC 113040

```

```

RESULT 29
AC162035
LOCUS      AC162035      200474 bp      DNA      linear      HTG 24-JUN-2005
DEFINITION Mus musculus chromosome 18 clone RP23-434N7 map 18, *** SEQUENCING
IN PROGRESS ***.
ACCESSION AC162035
VERSION   AC162035.4 GI:68163656
KEYWORDS  HTG; HTGS_PHRASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE    Mus musculus
ORGANISM  Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 200474)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.

```

Mus musculus chromosome 18, clone RP23-434N7

```

Unpublished
2 (bases 1 to 200474)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-MAY-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 200474)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JUN-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
On Jun 24, 2005 this sequence version replaced gi:67514128.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L32549
Center clone name: 434_N_7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and

```

\* the accession number will be preserved.  
\* 1 200474: contig of 200474 bp in length.

#### FEATURES

source  
Location/Qualifiers  
1. .200474  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="18"  
/map="18"  
/clone="RP23-434N7"  
/clone\_lib="RPCI-23 Female Mouse BAC"

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 200474;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTCC 25  
|||||  
Db 65699 AAAAAAGCAGACTGCTCTCC 65723

#### RESULT 30

AL845491

#### LOCUS

DEFINITION Mouse DNA sequence from clone RP23-360A2 on chromosome 2, complete sequence.

#### ACCESSION

AL845491

#### VERSION

AL845491.3 GI:23476758

#### KEYWORDS

HTG.

#### SOURCE

Mus musculus (house mouse)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 201381)

Direct Submission

Submitted (01-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Oct 2, 2002 this sequence version replaced gi:23337884.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; Sw.,

SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP

database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-360A2 is

from the RPCI-23 Mouse PAC library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

#### FEATURES

source

Location/Qualifiers

1. .201381

/organism="Mus musculus"

/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone="RP23-360A2"  
/clone\_lib="RPCI-23"

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 201381;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTCC 25  
|||||

Db 152299 AAAAAAGCAGCTGCTCTCC 152323

#### RESULT 31

AC165368

#### LOCUS

DEFINITION

AC165368

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jul 30, 2005 this sequence version replaced gi:70721574.

-----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site:<http://genome.wustl.edu>

-----

Center project name: M\_BA0028L24

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

-----

```
* 2700 2799: gap of unknown length
* 2800 4360: contig of 1561 bp in length
* 4361 4460: gap of unknown length
* 4461 5784: contig of 1324 bp in length
* 5785 7239: gap of unknown length
* 7240 7338: contig of 1354 bp in length
* 7339 8585: gap of unknown length
* 8586 8684: contig of 1246 bp in length
* 8685 10371: contig of 1687 bp in length
* 10372 10471: gap of unknown length
* 10472 12084: contig of 1613 bp in length
* 12085 12184: gap of unknown length
* 12185 14187: contig of 1903 bp in length
* 14188 14188: gap of unknown length
* 14189 15412: contig of 1225 bp in length
* 15413 15512: gap of unknown length
* 15513 16808: contig of 1295 bp in length
* 16809 16907: gap of unknown length
* 16908 19393: contig of 2486 bp in length
* 19394 19493: gap of unknown length
* 19494 21170: contig of 1677 bp in length
* 21171 21270: gap of unknown length
* 21271 22666: contig of 1396 bp in length
* 22667 22766: gap of unknown length
* 22767 24130: contig of 1364 bp in length
* 24131 24230: gap of unknown length
* 24231 25979: contig of 1749 bp in length
* 25980 26079: gap of unknown length
* 26080 27705: contig of 1626 bp in length
* 27706 27805: gap of unknown length
* 27806 29900: contig of 2095 bp in length
* 29901 30000: gap of unknown length
* 30001 33241: contig of 3241 bp in length
* 33242 33341: gap of unknown length
* 33343 36132: contig of 2791 bp in length
* 36133 36232: gap of unknown length
* 36233 39773: contig of 3541 bp in length
* 39774 39873: gap of unknown length
* 39874 42923: contig of 3056 bp in length
* 42924 43023: gap of unknown length
* 43024 46035: contig of 3006 bp in length
* 46036 46135: gap of unknown length
* 46136 49576: contig of 3441 bp in length
* 49577 49677: gap of unknown length
* 49678 52970: contig of 3294 bp in length
* 52971 53070: gap of unknown length
* 53071 56579: contig of 3509 bp in length
* 56580 60428: contig of 3749 bp in length
* 60429 60528: gap of unknown length
* 60529 64364: contig of 3836 bp in length
* 64365 64464: gap of unknown length
* 64465 69850: contig of 5386 bp in length
* 69851 69950: gap of unknown length
* 69951 74459: contig of 4509 bp in length
* 74460 74559: gap of unknown length
* 74560 78370: contig of 3811 bp in length
* 78371 78470: gap of unknown length
* 78471 83399: contig of 4929 bp in length
* 83400 83499: gap of unknown length
* 83500 89000: contig of 5501 bp in length
* 89001 89100: gap of unknown length
* 89101 94291: contig of 5191 bp in length
* 94292 94391: gap of unknown length
* 94392 99263: contig of 4872 bp in length
* 99264 99363: gap of unknown length
* 99364 104604: contig of 5241 bp in length
* 104605 104704: gap of unknown length
* 110375 110375: contig of 5670 bp in length
* 110375 110474: gap of unknown length
* 110475 115345: contig of 4871 bp in length
* 115346 115445: gap of unknown length
```

---

```
* 115446 120862: contig of 5417 bp in length
* 120863 120962: gap of unknown length
* 120963 126683: contig of 5721 bp in length
* 126684 126783: gap of unknown length
* 126784 132168: contig of 5585 bp in length
* 132169 132468: gap of unknown length
* 132469 139965: contig of 7497 bp in length
* 139966 140065: gap of unknown length
* 140066 147193: contig of 7128 bp in length
* 147194 147293: gap of unknown length
* 147294 153864: contig of 6571 bp in length
* 153865 153964: gap of unknown length
* 153965 165095: contig of 11131 bp in length
* 165096 165195: gap of unknown length
* 165196 176906: contig of 11711 bp in length
* 176907 177007: gap of unknown length
* 177008 185891: contig of 8885 bp in length
* 185892 185991: gap of unknown length
* 185992 201588: contig of 15597 bp in length
* 201589 201688: gap of unknown length
* 201689 203209: contig of 1521 bp in length.
* 203210 Location/Qualifiers
* 203211 1..203209
* 203212 /organism="Mus musculus"
* 203213 /mol_type="genomic DNA"
* 203214 /db_xref="taxon:10090"
* 203215 /chromosome="1"
* 203216 /clone="RP23-28L24"
* 203217 1..1276
* 203218 /note="assembly_name:Contig18"
* 203219 1277..1376
* 203220 /estimated_length=unknown
* 203221 1377..2699
* 203222 /note="assembly_name:Contig19"
* 203223 2700..2799
* 203224 /estimated_length=unknown
* 203225 2800..4360
* 203226 /note="assembly_name:Contig22"
* 203227 4361..4460
* 203228 /estimated_length=unknown
* 203229 4461..5784
* 203230 /note="assembly_name:Contig23"
* 203231 5785..5884
* 203232 /estimated_length=unknown
* 203233 5885..7238
* 203234 /note="assembly_name:Contig24"
* 203235 7239..7338
* 203236 /estimated_length=unknown
* 203237 7339..8584
* 203238 /note="assembly_name:Contig25"
* 203239 8585..8684
* 203240 /estimated_length=unknown
* 203241 8685..10371
* 203242 /note="assembly_name:Contig26"
* 203243 10372..10471
* 203244 /estimated_length=unknown
* 203245 10472..12084
* 203246 /note="assembly_name:Contig27"
* 203247 12085..12184
* 203248 /estimated_length=unknown
* 203249 12185..14087
* 203250 /note="assembly_name:Contig28"
* 203251 14088..14187
* 203252 /estimated_length=unknown
* 203253 14188..15412
* 203254 /note="assembly_name:Contig29"
* 203255 15413..15512
* 203256 /estimated_length=unknown
* 203257 15513..16807
* 203258 /note="assembly_name:Contig30"
* 203259 16808..16907
* 203260 /estimated_length=unknown
```

```

Query Match      87.2%; Score 21.8; DB 14; Length 203209;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTTCC 25
   |||||
Db 143973 AAAAAAGAGCGCAGACTGCTCTTCC 143956
   |||||

RESULT 33
AL591032 205573 bp DNA linear ROD 14-MAR-2002
LOCUS Mouse DNA sequence from clone RP22-211A10 on chromosome 4, complete
DEFINITION
ACCESSION AL591032
VERSION AL591032.19 GI:19571900
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205573)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hunquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi:18476682.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; Swi;
SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP22-211A10 is
from the RPCI-22 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP22-211A10.

FEATURES
source
1..205573
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosomes="4"
   /clone="RP22-211A10"
   /clone_lib="RPCI-22"
misc_feature
180212
   /note="Tandem repeat. Forced join. Gap size estimated to
be approximately 160bp by restriction digest data."

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 205573;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTTCC 25
   |||||
Db 189760 AAAAAAGAGCGCAGACTGCTCTTCC 189784
   |||||

RESULT 34

```

```

Query Match      87.2%; Score 21.8; DB 14; Length 203209;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTTCC 25
   |||||
Db 194973 AAAAAAGAGCGCAGACTGCTCTTCC 194997
   |||||

RESULT 32
CR848808 205459 bp DNA linear HTG 03-AUG-2005
LOCUS Mus musculus chromosome 2 clone RP23-442I10, WORKING DRAFT
DEFINITION
ACCESSION CR848808
VERSION CR848808.9 GI:71794035
KEYWORDS HTG; HTGS PHASE2; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205459)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 3, 2005 this sequence version replaced gi:68645812.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk
----- Project Information
Center project name: bm442110
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 201342 bases at least Q40
Consensus quality: 201666 bases at least Q30
Consensus quality: 201997 bases at least Q20
Insert size: 205459; sum-of-contigs
Insert size: 142269; 3.2% error; agarose-fp
Quality coverage: 11.00x in Q20 bases; sum-of-contigs Quality
coverage: 16.14x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 205459: contig of 205459 bp in length.
Location/Qualifiers
1..205459
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosomes="2"
   /clone="RP23-442I10"
   /clone_lib="RPCI-23"
misc_feature
1..205459
   /note="assembly_fragment:01572
clone_end:SP6
vector_side:left"

ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 205459;
Best Local Similarity 92.0%; Pred. No. 12;

```

AC027767/c  
LOCUS  
DEFINITION

AC027767 206224 bp DNA linear HTG 14-JUN-2003  
Mus musculus chromosome 11 clone RP23-51L13 map 11, WORKING DRAFT  
SEQUENCE, 5 unordered pieces.

AC027767  
AC027767.4 GI:31745253  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 206224)  
Birren,B., Nusbaum,C. and Lander,E.

Mus musculus chromosome 11, clone RP23-51L13  
Unpublished  
2 (bases 1 to 206224)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRoque,K., Lanazares,R., Landers,T., Lehoczyk,J.,  
Levine,R., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mieng,J., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,K., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Vieler,J., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 206224)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cooke,A., Cooke,P., Corum,B., DeArelano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mieng,J., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tessaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Vieler,J., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 14, 2003 this sequence version replaced gi:14547878.  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green,P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITB  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8885

Center clone name: 51.L13

----- Summary Statistics

Sequencing vector: M13; M77815; 29% of reads

Sequencing vector: Plasmid; n/a; 71% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 205175 bases at least Q40

Consensus quality: 205524 bases at least Q30

Consensus quality: 205691 bases at least Q20

Insert size: 203000; agarose-fp

Insert size: 205824; sum-of-contigs

Quality coverage: 14.6 in Q20 bases; agarose-fp

Quality coverage: 14.4 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 45102: contig of 45102 bp in length

\* 45103 45202: gap of 100 bp

\* 45203 146534: contig of 101332 bp in length

\* 146535 146634: gap of 100 bp

\* 146635 159842: contig of 13208 bp in length

\* 159843 159942: gap of 100 bp

\* 159943 201592: contig of 41650 bp in length

\* 201593 201692: gap of 100 bp

\* 201693 206224: contig of 4532 bp in length.

FEATURES

source

1..206224

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="11"

/map="11"

/clone="RP23-51L13"

/clone\_lib="RP23-51L13 Female Mouse BAC"

1..45102

/notes="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

45103..45202

/estimated\_length=100

45203..146534

/notes="assembly\_fragment"

146535..146634

/estimated\_length=100

146635..159842

/notes="assembly\_fragment"

159843..159942

/estimated\_length=100

159943..201592

/notes="assembly\_fragment"

201593..201692

/estimated\_length=100

201693..206224

/notes="assembly\_fragment"

clone\_end:T7

vector\_side:right"

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 206224;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCGACTGCTCTCC 25

Db 185703 AAAAAAGCGCGACTGCTCTCC 185679





/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/clone="RP23-83023"	
/clone_lib="RPC1 mouse BAC library 23"	
gap	1126..1225
gap	/estimated_length=unknown
gap	2422..2521
gap	/estimated_length=unknown
gap	4798..4897
gap	/estimated_length=unknown
gap	6027..6125
gap	/estimated_length=unknown
gap	7281..7380
gap	/estimated_length=unknown
gap	10171..10270
gap	/estimated_length=unknown
gap	26392..26491
gap	/estimated_length=unknown
gap	44421..44520
gap	/estimated_length=unknown
gap	66707..66806
gap	/estimated_length=unknown
gap	95201..95300
gap	/estimated_length=unknown
gap	124528..124627
gap	/estimated_length=unknown
gap	156742..156841
gap	/estimated_length=unknown
gap	187801..187900
gap	/estimated_length=unknown
ORIGIN	
Query Match	87.2%; Score 21.8; DB 14; Length 208910;
Best Local Similarity	92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 AAAAAAGAGCGAGCTGCTTCC 25
Db	148851 AAAAAAGAGCTGCTTCC 148875
RESULT 37	
AC102988	
LOCUS	
DEFINITION	AC102988 213321 bp DNA linear HTG 09-NOV-2002
	Rattus norvegicus clone CH230-60E23, WORKING DRAFT SEQUENCE, 2
	unordered pieces.
ACCESSION	AC102988
VERSION	AC102988.5 GI:24819571
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 213321)
	Muzny, D., Marle, J., Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bliswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Delgado, O., Denson, S., Davy-Carroll, L., De Anda, C., Dederich, D., Devila, M.L., Davis, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Evans, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Haravay, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J.,

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 112868: contig of 112868 bp in length  
 \* 112869 112968: gap of unknown length  
 \* 112969 213321: contig of 100353 bp in length.

## FEATURES

source

1. .213321  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"

misc\_feature

/clone="CH230-60E23"  
 complement(95870..96694)  
 /note="clone boundary"

clone end:Sp6

site:EcoRI

end\_sequence:BH364717"

112869..112968

/estimated\_length=unknown

211383..213321

/note="wgs\_contig"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 213321;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 AAAAAAGAGCGAGCTGCTCTCC 25

|||||  
 Db 8548 AAAAAAGAGCAGACTGCTCTCC 8572

RESULT 38

AL845494

LOCUS

AL845494 216517 bp DNA linear ROD 09-OCT-2002  
 Mouse DNA sequence from clone RP23-330D3 on chromosome 2, complete  
 sequence.

ACCESSION

AL845494

VERSION

AL845494.3 GI:23895988

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 216517)

REFERENCE

Tracey,A.

AUTHORS

Direct Submission

TITLE

Submitted (09-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

On Oct 12, 2002 this sequence version replaced gi:23337886.

COMMENT

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Swi.,  
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP  
 database can be found at

[http://www.sanger.ac.uk/Projects/c\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/c_elegans/wormpep) RP23-330D3 is  
 from the RPCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

## FEATURES

source

1. .216517

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="2"

/clone="RP23-330D3"

/clone\_lib="RPCI-23"

## ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 216517;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 AAAAAAGAGCGAGCTGCTCTCC 25

|||||

Db 186644 AAAAAAGAGCAGCTGCTCTCC 186668

RESULT 39

BX813319/c

LOCUS

BX813319 217197 bp DNA linear ROD 06-FEB-2004  
 Mouse DNA sequence from clone RP23-54B12 on chromosome 2, complete  
 sequence.

ACCESSION

BX813319

VERSION

BX813319.4 GI:42414974

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 217197)

REFERENCE

Sycamore,N

AUTHORS

Direct Submission

TITLE

Submitted (06-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

On Feb 5, 2004 this sequence version replaced gi:41060267.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one cDNA subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep/RP23-54B12](http://www.sanger.ac.uk/projects/C_elegans/wormpep/RP23-54B12) is

from the RPCI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

#### FEATURES

Source

1. 217197

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="2"

/clone="RP23-54B12"

/clone\_lib="RPCI-23"

#### ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 217197;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTTCC 25

||||| | | | | | | | | | | | | | | | | | | | |

Db 91248 AAAAAAGAGCAGCTGCTTCC 91224

#### RESULT 40

AC095149

LOCUS

AC095149 Rattus norvegicus clone CH230-813, WORKING DRAFT SEQUENCE, 3

DEFINITION

unordered pieces.

AC095149

AC095149.6 GI:30467261

HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Muridae; Murinae; Rattus.

1 (bases 1 to 217867)

Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Amin,A., Anguitano,D.,

Allen,C., Allen,H., Alsbrooks,S., Ayodeji,M., Baca,E., Baden,H.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Barber,M., Barnstead,M., Benahmed,F.,

Baldwin,D., Bandaranaike,D., Blankenburg,K., Blyth,P., Brown,M.,

Biswal,K., Blair,J., Burch,P., Burrell,K., Calderon,E.,

Bryant,N., Buhay,C., Carter,K., Cavazos,I., Cesar,H., Center,A.,

Cardenas,V., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Chacko,J., Chaveze,D., Chen,G., Coyle,M., Cree,A., D'Souza,L.,

Cleveland,C., Cockrell,R., Cox,C., Dalrymple,L., De Anda,C., Dederich,D.,

Devila,M.L., Davis,C., Davy-Carroll,L., Ding,Y., Dinh,H., Divya,K.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Durbin,K., Duval,B., Eaves,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Evans,C.A., Falls,T., Fan,G.,

Egan,A., Escotto,M., Eugene,C., Garcia,A., Garner,T., Foster,M., Foster,P.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Garza,M.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebrgeorgis,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,J., London,P., Longacre,S., Lopez,J.,

Lorensuhewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaokemele,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,  
Pascernak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L.,  
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Reilly,M., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
Williams,G., Wilson,R., Wlaczek,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

#### Direct Submission

Unpublished

2 (bases 1 to 217867)

Worley,K.C.

#### Direct Submission

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GCTH

Center clone name: CH230-813

Summary Statistics

Assembly program: Atlas

Consensus quality: 205299 bases at least Q40

Consensus quality: 207638 bases at least Q30

Consensus quality: 208930 bases at least Q20

Estimated insert size: 216123; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.









```

RESULT 44
BX890623      226542 bp   DNA   linear   ROD 26-FEB-2004
LOCUS
DEFINITION
  Mouse DNA sequence from clone RP23-83023 on chromosome 2, complete
  sequence.
ACCESSION
  BX890623
VERSION
  BX890623.6   GI:43823581
KEYWORDS
  HTG.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 226542)
REFERENCE
  1 (bases 1 to 226542)
AUTHORS
  Direct Submission
  Submitted (26-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
  On Feb 26, 2004 this sequence version replaced gi:42627462.
  Sequence from the Mouse Genome Sequencing Consortium whole genome
  shotgun may have been used to confirm this sequence. Sequence data
  from the whole genome shotgun alone has only been used where it has
  a phred quality of at least 30.
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  -----
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  This sequence was finished as follows unless otherwise noted: all
  regions were either double-stranded or sequenced with an alternate
  chemistry or covered by high quality data (i.e. phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by at least
  one plasmid subclone or more than one M13 subclone; and the
  assembly was confirmed by restriction digest, except on the rare
  occasion of the clone being a YAC.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
  on the WORMPEP database can be found at
  http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-83023 is
  from the RPCI-23 Mouse BAC Library
  constructed by the group of Pieter de Jong.
  For further details see http://www.chori.org/bacpac/home.htm
  VECTOR: pBACe3.6.
  Location/Qualifiers
    1..226542
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /chromosome="2"
      /clones="RP23-83023"
      /clone_lib="RPCI-23"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 226542;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGCTGCTTCC 25
    |||||
Db 41228 AAAAAAGCAGCTGCTTCC 41252
    |||||

RESULT 45

```

```

AC128603/c
LOCUS
DEFINITION
  Rattus norvegicus clone CH230-227D15, *** SEQUENCING IN PROGRESS
  ***. 4 unordered pieces.
ACCESSION
  AC128603
VERSION
  AC128603.2   GI:23196137
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridae; Murinae; Rattus.
  1 (bases 1 to 229093)
REFERENCE
  1 (bases 1 to 229093)
AUTHORS
  Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
  Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
  Anylebechin, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
  Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Ben.,
  Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
  Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
  Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
  Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
  Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
  Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
  Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
  Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
  Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
  Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
  Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
  Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
  Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
  Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
  Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
  Hollings, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
  Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
  Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
  Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
  Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
  Loreschwah, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J.,
  Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
  Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
  Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
  Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
  Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
  Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
  Nwaokemele, O., Okwuon, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
  Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
  Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
  Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
  Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
  Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
  Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
  Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
  Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,
  Steinkle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
  Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
  Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
  Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
  Williams, G., Willson, R., Wlaczek, R., Wooden, H., Worley, K.,
  Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
  Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
  Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
  Weinstein, G. and Gibbs, R.A.
  Direct Submission
  Unpublished
  2 (bases 1 to 229093)
  Worley, K.C.
  Direct Submission
  Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 229093)
  Rat Genome Sequencing Consortium.
  Direct Submission
  TITLE
  JOURNAL
  REFERENCE
  AUTHORS
  TITLE
  JOURNAL
  REFERENCE
  AUTHORS
  TITLE

```





of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23265039.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
-----  
Center project name: GY40  
Center clone name: CH230-154H6  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 220971 bases at least Q40  
Consensus quality: 224061 bases at least Q30  
Consensus quality: 226217 bases at least Q20  
Estimated insert size: 230107; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 97368: contig of 97368 bp in length  
\* 97369: gap of unknown length  
\* 97459: contig of 136063 bp in length  
\* 233531: gap of unknown length  
\* 233632: contig of 1750 bp in length.

## FEATURES

source  
1. .235381  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-154H6"  
1. .2102  
/note="wgs end extension  
clone\_end:Sp6"  
4409. .5070  
/note="clone\_boundary  
clone\_end:Sp6  
site:  
end sequence:BH362740"  
97369. 97468  
/estimated\_length=unknown  
129461. .129613  
/note="clone\_boundary  
clone\_end:T7  
site:  
end sequence:BH362739"  
230457. .231609  
/note="wgs\_end extension  
clone\_end:T7"  
231949. .233531  
/note="wgs\_end extension  
clone\_end:T7"

233532. .233631  
/estimated\_length=unknown

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 235381;  
Best Local Similarity 92.0%; Pred. No. 12;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTGCTCTTCC 25  
|||||  
45019 AAAAAGAGCGTGAATGCTCTTCC 45043  
|||||

RESULT 47  
AC103495/c

LOCUS  
DEFINITION  
AC103495 237321 bp DNA linear HTG 10-MAY-2003  
Rattus norvegicus clone CH230-31E1, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
4 unordered pieces.

AC103495  
AC103495.7 GI:30520399  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS  
Rattus norvegicus (Norway rat)

SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE  
1 (bases 1 to 237321)  
Muzny, D., Maric, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Evans, C., Evans, C., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M.,  
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, V., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S., Hodge, A., Hoques, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshewa, L., Louseged, H., Lozado, R., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M., McNeill, T., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Sitter, C., Smajs, D.,  
Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasa, D., Waldron, L., Walker, B., Wang, J.,  
Wang, O., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, C. and Gibbs, R. A.

## Direct Submission

Unpublished

2 (bases 1 to 237321)

Worley, K.C.

## Direct Submission

Submitted (25-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 237321)

Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24635679.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GHTM

Center clone name: CH230-31E1

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 220013 bases at least Q40

Consensus quality: 222844 bases at least Q30

Consensus quality: 224875 bases at least Q20

Estimated insert size: 231039; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 232262: contig of 232262 bp in length

\* 232263 232362: gap of unknown length

\* 232363 234000: contig of 1638 bp in length

\* 234001 234100: gap of unknown length

\* 234101 235489: contig of 1389 bp in length

\* 235490 235589: gap of unknown length

\* 235590 237321: contig of 1732 bp in length.

Location/Qualifiers

1. .237321

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-31E1"

1. .1834

/note="wgs end extension

clone end:Sp6"

4516. .5301

/note="clone boundary  
clone end:Sp6  
site:EcoRI  
end\_sequence:BH320543"  
complement(124058. .231882)  
/note="clone boundary  
clone end:T7  
site:EcoRI  
end\_sequence:BH320542"  
129379. .131737  
/note="wgs end extension  
clone end:T7"  
232263. .232362  
/estimated\_length=unknown  
234001. .234100  
/estimated\_length=unknown  
235490. .235589  
/estimated\_length=unknown

## misc\_feature

## misc\_feature

## gap

## gap

## gap

## ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 237321;

Best Local Similarity 92.0%; Pred. NO. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25

Db 115298 AAAAAAGAGCGCTGACTGCTTCC 115274

## RESULT 48

AC126306

AC126306

LOCUS

DEFINITION

\*\*\*, 6 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

1 (bases 1 to 238355)

REFERENCE

AUTHORS

Murzyn, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Louisaeged, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwankwem, O., Okwuonu, G., Olarnunagsoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, F., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., Von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G. and Gibbs, R.A.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Unpublished  
 2 (bases 1 to 238355)  
 Worley, K.C.  
 Direct Submission  
 Submitted (05-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

3 (bases 1 to 238355)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (09-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 19, 2002 this sequence version replaced gi:21699124.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.

## COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GZIV  
 Center clone name: CH230-207N13  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 185501 bases at least Q40  
 Consensus quality: 188832 bases at least Q30  
 Consensus quality: 190590 bases at least Q20  
 Estimated insert size: 233967; sum-of-contigs estimation  
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 211754: contig of 211754 bp in length  
 \* 211755 211854: gap of unknown length  
 \* 211855 228024: contig of 16170 bp in length  
 \* 228025 228124: gap of unknown length  
 \* 228125 229182: contig of 1058 bp in length  
 \* 229183 229282: gap of unknown length

\* 229283 230367: contig of 1085 bp in length  
 \* 230368 230467: gap of unknown length  
 \* 230468 232860: contig of 2393 bp in length  
 \* 232861 232960: gap of unknown length  
 \* 232961 238355: contig of 5395 bp in length.  
 FEATURES  
 source  
 1..238355  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-207N13"  
 misc\_feature  
 1..1097  
 /note="wgs\_contig"  
 misc\_feature  
 26398..29670  
 /note="wgs\_contig"  
 gap  
 211755..211854  
 /estimated\_length=unknown  
 misc\_feature  
 211855..213065  
 /note="wgs\_contig"  
 gap  
 228025..228124  
 /estimated\_length=unknown  
 gap  
 229183..229282  
 /estimated\_length=unknown  
 gap  
 230368..230467  
 /estimated\_length=unknown  
 gap  
 232861..232960  
 /estimated\_length=unknown  
 ORIGIN  
 Query Match 87.2%; Score 21.8; DB 14; Length 238355;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAGCGGAGCTGCTCTCC 25  
 Db 226727 AAAAAAGCGGAGCTGCTCTCC 226751  
 RESULT 49  
 AC128432/c  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-103P9, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 13 unordered pieces.  
 AC128432  
 AC128432.3 GI:25007946  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT: HTGS\_ENRICHED.  
 KEYWORDS  
 Rattus norvegicus (Norway rat)  
 SOURCE  
 Rattus norvegicus  
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus.  
 ORGANISM  
 1 (bases 1 to 239591)  
 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, N., Hernandez, J.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jacob, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisseon,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 239591)  
Worley,K.C.

Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 239591)  
Baylor College of Medicine  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GYXL  
Center clone name: CH230-103P9  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 175830 bases at least Q40  
Consensus quality: 180468 bases at least Q30  
Consensus quality: 183438 bases at least Q20  
Estimated insert size: 175856; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 130495: contig of 130495 bp in length  
\* 130496 130595: gap of unknown length  
\* 130596 13576: contig of 4981 bp in length  
\* 135767 13576: gap of unknown length  
\* 135777 216234: contig of 80558 bp in length  
\* 216235 216334: gap of unknown length  
\* 216335 220030: contig of 3696 bp in length  
\* 220031 220130: gap of unknown length  
\* 220131 221542: contig of 1412 bp in length  
\* 221543 221642: gap of unknown length  
\* 221643 222799: contig of 1157 bp in length  
\* 222800 222899: gap of unknown length  
\* 222900 224107: contig of 1208 bp in length  
\* 224108 225254: contig of 1047 bp in length  
\* 225255 225354: gap of unknown length  
\* 225355 226607: contig of 1253 bp in length  
\* 226608 226707: gap of unknown length  
\* 226708 228412: contig of 1705 bp in length  
\* 228413 228512: gap of unknown length  
\* 228513 230106: contig of 1594 bp in length  
\* 230107 230207: gap of unknown length  
\* 230207 233662: contig of 3456 bp in length  
\* 233663 239591: contig of 5829 bp in length.  
\* 23763 Location/Qualifiers  
1. .239591  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-103P9"  
10911..13043  
/note="wgs contig"  
22365..25174  
/note="wgs contig"  
62809..64554  
/note="wgs contig"  
130496..130595  
/estimated\_length=unknown  
135577..135676  
/estimated\_length=unknown  
159696..160700  
/note="wgs contig"  
189547..190698  
/note="wgs contig"  
216235..216334  
/estimated\_length=unknown  
220031..220130  
/estimated\_length=unknown  
221543..221642  
/estimated\_length=unknown  
222800..222899  
/estimated\_length=unknown  
224108..224207  
/estimated\_length=unknown  
225255..225354  
/estimated\_length=unknown  
226608..226707  
/estimated\_length=unknown  
228413..228512  
/estimated\_length=unknown  
230107..230206  
/estimated\_length=unknown  
233663..233762  
/estimated\_length=unknown

FEATURES  
source  
misc\_feature  
misc\_feature  
misc\_feature  
gap  
gap  
misc\_feature  
misc\_feature  
gap  
gap  
gap  
gap  
gap  
gap  
gap  
gap  
gap

**ORIGIN**  
 Query Match 87.2%; Score 21.8; DB 14; Length 239591;  
 Best Local Similarity 92.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
**Qy** 1 AAAAAAGAGCGGACTGCTCTCC 25  
 |||||  
**Db** 198883 AAAAAAGAGCGGACTGCTCTCC 198859  
  
**RESULT 50**  
**AC103505/c**  
**LOCUS**  
**DEFINITION** AC103505 239946 bp DNA linear HTG 10-MAY-2003  
 Rattus norvegicus clone CH230-94F6, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 2 unordered pieces.  
**ACCESSION** AC103505 GI:30520584  
**VERSION** HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
**KEYWORDS** Rattus norvegicus (Norway rat)  
**SOURCE** Rattus norvegicus  
**ORGANISM** Rattus norvegicus  
  
**REFERENCE**  
**AUTHORS** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Murioidea; Muridae; Murinae; Rattus.  
 Muzny, D., Marie, C., Allen, H., Metzker, M., Lee, A., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, N., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, C., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gueorge, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshewari, M., Louisedge, H., Lozado, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Napua, P., Martin, K., Martin, R., Martinez, E.,  
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwaokemele, O., Okunolu, G., Olarnpungagoon, A., Pal, S., Parks, K.,  
 Paternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C.,  
 Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,  
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, J., Zhou, D., von  
 Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O.,

```
misc_feature      complement(235421..236183)
                  /note="clone_boundary
                  clone_end:T7
                  site:ECORI
                  end_sequence:BH330539"
gap               237950..238049
                  /estimated_length=unknown

ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 239946;
Best Local Similarity 92.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAAAAAGAGCGGAGACTGCTCTTCC 25
        |||||
Db      41607 AAAAAAGAGCGGAGCTGCTCTTCC 41583
```

Search completed: February 3, 2006, 21:27:55  
Job time : 892 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 10:53:33 ; Search time 203.111 Seconds  
(without alignments)  
820.326 Million cell updates/sec

Title: US-10-719-900-35

Perfect score: 25

Sequence: 1 aaaaagagcgagactggtcttcc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq\_21.\*

1: Geneseqn1980s.\*

2: Geneseqn1980s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.8	87.2	36599	13	ABD33506
2	20.2	80.8	1650	13	ABD48813
3	19.8	79.2	68842	14	ADZ13146
4	19.4	77.6	30752	11	ACN44344
5	19.4	77.6	90091	10	ADC85509
6	19.4	77.6	90100	9	ADA03029
7	19.4	77.6	90100	10	ADB72767
8	19.4	77.6	90100	12	ADM74624
9	19.2	76.8	1276	8	AAD55112
10	19.2	76.8	1276	8	ACC45096
11	19.2	76.8	1392	9	ACA36353
12	19.2	76.8	29956	9	ADA02963
13	19.2	76.8	29956	10	ADB72701
14	19.2	76.8	29956	10	ADC85443
15	19.2	76.8	29956	12	ADM74558
16	19.2	76.8	110000	6	ABA03041_24
17	19.2	76.8	123192	13	ADV34995
18	18.8	75.2	43445	12	ADH18345
19	18.8	75.2	43445	12	ADO32886

20	18.6	74.4	39405	11	ACN44704	Acn44704	Mouse gen
21	18.6	74.4	39790	11	ACN44796	Acn44796	Mouse gen
C 22	18.6	74.4	47115	9	ADA02627	Ada02627	Mouse Flt
C 23	18.6	74.4	47115	10	ADB72365	Adb72365	Mouse Flt
C 24	18.6	74.4	47115	10	ADe95875	Ad95875	Mouse Flt
C 25	18.6	74.4	49281	11	ACN44048	Acn44048	Human can
C 26	18.6	74.4	64482	12	ADQ59515	Adq59515	Human can
C 27	18.6	74.4	64482	14	ADZ13896	Adz13896	Murine ca
C 28	18.6	74.4	89213	13	ABD33614	Abd33614	Murine ca
C 29	18.6	74.4	92099	14	ADZ12547	Adz12547	Murine ca
C 30	18.4	73.6	47115	9	ADA02627	Ada02627	Mouse Flt
31	18.4	73.6	47115	10	ADB72365	Adb72365	Mouse Flt
32	18.4	73.6	47115	10	ADe95875	Ad95875	Mouse Flt
33	18.4	73.6	92099	14	ADZ12547	Adz12547	Murine ca
C 34	18.2	72.8	337	8	ABZ55378	Abz55378	Aspergill
C 35	18.2	72.8	92861	12	ADQ97843	Adq97843	Mouse can
C 36	17.8	71.2	95	14	ADX04494	Adx04494	Rat prima
37	17.8	71.2	331	6	ABSG9109	Absg9109	Novel mur
38	17.8	71.2	3486	12	ADO36057	Ado36057	Novel mou
C 39	17.8	71.2	3486	12	ADO36057	Ado36057	Novel mou
40	17.8	71.2	4333	10	ADB81347	Adb81347	Murine do
41	17.8	71.2	4333	12	ADQ76467	Adq76467	Nucleotid
42	17.8	71.2	4333	13	ADQ82251	Adq82251	Rosa26 lo
43	17.8	71.2	4333	13	ADR04031	Adr04031	Murine Ro
44	17.8	71.2	4333	13	ADT71363	Adt71363	3' arm Ro
45	17.8	71.2	5409	12	ADN11341	Adn11341	Murine Ro
46	17.8	71.2	10491	10	ADB81348	Adb81348	PROSA12 v
47	17.8	71.2	11176	4	AAS01080	Aas01080	Mouse Fos
48	17.8	71.2	11176	6	ADT22528	Adt22528	Mouse Fos
49	17.8	71.2	11784	6	ABT08201	Abt08201	Modified
50	17.8	71.2	11784	10	ADB81350	Adb81350	Bacteriop
51	17.8	71.2	12538	10	ADB81342	Adb81342	Targeting
52	17.8	71.2	12607	13	ADQ82912	Adq82912	Vector co
53	17.8	71.2	12645	10	ADB81343	Adb81343	Targeting
54	17.8	71.2	12739	13	ADB04033	Adb04033	Murine Ro
55	17.8	71.2	12739	13	ADT71365	Adt71365	Rosa26 ta
56	17.8	71.2	13139	12	ADN11340	Adn11340	Murine Ro
57	17.8	71.2	13139	12	ADQ76465	Adq76465	Nucleotid
58	17.8	71.2	13139	13	ADQ82249	Adq82249	Rosa26 lo
59	17.8	71.2	13139	13	ADR04029	Adr04029	Murine Ro
60	17.8	71.2	13139	13	ADT71361	Adt71361	Mouse Ros
61	17.8	71.2	13544	13	ADR04032	Adr04032	Murine Ro
62	17.8	71.2	13544	13	ADT71364	Adt71364	Rosa26 ta
63	17.8	71.2	14411	12	ADQ76469	Adq76469	Nucleotid
64	17.8	71.2	14411	13	ADQ82253	Adq82253	Targeting
65	17.8	71.2	14947	12	ADN11343	Adn11343	Targeting
66	17.8	71.2	15174	12	ADN11347	Adn11347	Targeting
67	17.8	71.2	15199	12	ADN11345	Adn11345	Targeting
68	17.8	71.2	15342	13	ADR04034	Adr04034	Murine Ro
69	17.8	71.2	15342	13	ADT71366	Adt71366	Rosa26 lo
70	17.8	71.2	18352	12	ADG44924	Adg44924	Mouse cat
71	17.8	71.2	18666	13	ADQ82254	Adq82254	FABP-Cre
C 72	17.8	71.2	21635	11	ACN44628	Acn44628	Mouse gen
73	17.8	71.2	21730	13	ADQ82255	Adq82255	Cre subst
74	17.8	71.2	22114	13	ABD33131	Abd33131	Murine ca
75	17.8	71.2	25543	11	ACN44384	Acn44384	Mouse gen
76	17.8	71.2	26666	11	ACN45032	Acn45032	Mouse gen
C 77	17.8	71.2	30654	11	ACN45204	Acn45204	Mouse gen
C 78	17.8	71.2	32847	12	ADQ97767	Adq97767	Mouse can
79	17.8	71.2	35183	13	ABD33355	Abd33355	Murine ca
80	17.8	71.2	35359	11	ACN44128	Acn44128	Mouse gen
C 81	17.8	71.2	43591	11	ACN44916	Acn44916	Mouse gen
82	17.8	71.2	44972	11	ACN44080	Acn44080	Mouse gen
83	17.8	71.2	47299	13	ABD33226	Abd33226	Murine ca
C 84	17.8	71.2	49088	11	ACN43856	Acn43856	Mouse gen
85	17.8	71.2	49349	13	ADRe6997	Adre6997	Mouse can
C 86	17.8	71.2	49349	13	ADRe6997	Adre6997	Mouse can
87	17.8	71.2	49349	14	ADZ12842	Adz12842	Murine ca
C 88	17.8	71.2	49349	14	ADZ12842	Adz12842	Murine ca
C 89	17.8	71.2	54303	13	ACN44340	Acn44340	Mouse gen
90	17.8	71.2	54303	13	ABD32752	Abd32752	Mouse gen
91	17.8	71.2	55544	11	ACN44388	Acn44388	Mouse gen
C 92	17.8	71.2	57860	12	ADQ97692	Adq97692	Mouse can

93	17.8	71.2	58665	11	ACN45164	Acn45164 Mouse gen	KW	Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
94	17.8	71.2	65559	11	ACN44408	Acn44408 Mouse gen	KW	ds; cancer; cytostatic.
95	17.8	71.2	65559	11	ACN44408	Acn44408 Mouse gen	XX	
- c 96	17.8	71.2	65952	12	AD059527	Adq59527 Human can	OS	Mus musculus.
c 97	17.8	71.2	66009	13	AD213908	Adz13908 Murine ca	XX	
c 98	17.8	71.2	66126	14	AD333341	Abd33341 Murine ca	PN	WO2004058146-A2.
c 99	17.8	71.2	69515	11	ACN44660	Acn44660 Mouse gen	XX	
c 100	17.8	71.2	70215	11	ACN43992	Acn43992 Mouse gen	PD	15-JUL-2004.
c 101	17.8	71.2	71594	12	ADQ97761	Adq97761 Mouse can	XX	
c 102	17.8	71.2	75782	12	ADQ97795	Adq97795 Mouse can	PF	15-DEC-2003; 2003WO-US040081.
c 103	17.8	71.2	77530	13	ABD33240	Abd33240 Murine ca	XX	
c 104	17.8	71.2	79256	13	ABD33177	Abd33177 Murine ca	PR	17-DEC-2002; 2002US-00322281.
c 105	17.8	71.2	80423	13	ABD32576	Abd32576 Mouse can	XX	
c 106	17.8	71.2	84410	13	ABD33551	Abd33551 Murine ca	PA	(SAGR-) SAGRES DISCOVERY INC.
c 107	17.8	71.2	88493	13	ABD33522	Abd33522 Murine ca	XX	
c 108	17.8	71.2	92969	12	ADQ97202	Adq97202 Mouse can	PI	Morris DW, Malandro MS;
c 109	17.8	71.2	96389	9	ADA02675	Ada02675 Mouse Top	XX	WPI; 2004-499109/47.
c 110	17.8	71.2	96389	10	ADB72413	Adb72413 Mouse Top	DR	
c 111	17.8	71.2	96389	10	ADQ95923	Adq95923 Mouse Top	XX	
c 112	17.8	71.2	96599	9	ADA02933	Ada02933 Mouse Bra	PT	Novel human cancer associated protein encoded within open reading frame
c 113	17.8	71.2	96599	10	ADB72671	Adb72671 Mouse Bra	PT	of cancer associated gene, useful as targets for diagnosing cancer.
c 114	17.8	71.2	96599	10	ADC85413	Adc85413 Mouse Bra	PS	Disclosure; SEQ ID NO 679; 182pp; English.
c 115	17.8	71.2	96599	12	ADM74528	Adm74528 Murine ca	XX	
c 116	17.8	71.2	103190	14	AD213270	Adz13270 Murine ca	XX	
c 117	17.8	71.2	103661	12	ADQ97712	Adq97712 Mouse can	CC	The invention relates to cancer-associated proteins (CAP) and the cancer-
c 118	17.8	71.2	110000	8	ABX16390_1	Continuation (2 of	CC	associated (CA) nucleic acids encoding them. The invention also relates
c 119	17.8	71.2	110000	11	ACN43984_1	Continuation (2 of	CC	to a method for treating cancers involving administering to a patient an
c 120	17.8	71.2	111252	11	ACN44444	Acn44444 Mouse gen	CC	inhibitor of CAP, and a method of screening for anticancer activity in a
c 121	17.8	71.2	130427	13	ADV35011	Adv35011 Murine cd	CC	potential drug involving providing a cell that expresses a CA gene,
c 122	17.8	71.2	150830	12	ADQ97260	Adq97260 Mouse can	CC	contacting a tissue sample derived from a cancer cell with an anticancer
c 123	17.8	71.2	171936	6	ABE56565	ABe56565 Human SUL	CC	drug candidate and monitoring the effect of the anticancer drug candidate
c 124	17.8	71.2	171936	12	ADN16205	Adn16205 Mouse sul	CC	on expression of the CA gene. The CAP proteins are useful for detecting
c 125	17.8	71.2	215126	12	ADQ97362	Adq97362 Mouse can	CC	cancer associated with expression of a CAP protein in a test cell sample
c 126	17.8	71.2	215980	6	AAI38337	AAI38337 Complemen	CC	and for screening for a bioactive agent capable of modulating the
c 127	17.8	71.2	227448	13	ABD32841	Abd32841 Mouse can	CC	activity of a CAP protein. The CA nucleic acids are useful for diagnosing
c 128	17.8	71.2	241748	14	AD213116	Adz13116 Murine ca	CC	cancer, involving determining the expression of a CA nucleic acid in a
c 129	17.8	71.2	241748	14	AD213116	Adz13116 Murine ca	CC	tissue. This sequence represents a murine CA gene of the invention. Note:
c 130	17.6	70.4	459	6	ABE69484	ABe69184 Novel mur	CC	The sequence data for this patent did not form part of the printed
c 131	17.6	70.4	1285	2	AAZ34328	Aaz34328 Human EST	CC	specification, but was obtained in electronic format directly from WIPO
c 132	17.6	70.4	1285	3	AACT78594	Aac78594 Human EST	CC	at ftp.wipo.int/pub/published_pct_sequences
c 133	17.6	70.4	1285	8	ACA63896	ACA63896 Novel hum	XX	
c 134	17.6	70.4	1285	8	ACA72060	ACA72060 Human PRO	XX	Sequence 36599 BP; 10225 A; 7259 C; 8034 G; 10847 T; 0 U; 234 Other;
c 135	17.6	70.4	1285	8	ABX92700	ABx92700 Human PRO		
c 136	17.6	70.4	1285	8	ACA66441	ACA66441 Human sec		Query Match 87.2%; Score 21.8; DB 13; Length 36599;
c 137	17.6	70.4	1285	9	ADA25067	Ada25067 Novel hum		Best Local Similarity 92.0%; Pred. No. 4.5;
c 138	17.6	70.4	1285	9	ACD30042	AcD30042 Novel hum		Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
c 139	17.6	70.4	1285	9	ADA12728	Ada12728 Human sec		
c 140	17.6	70.4	1285	9	ACD29457	AcD29457 Novel hum	QY	1 AAAAAAGAGCGCAGACTGCTCTTCC 25
c 141	17.6	70.4	1285	10	ADB74034	Adb74034 Human PRO	Db	35597 AAAAAAGAGCAGACTGCTCTTCC 35621
c 142	17.6	70.4	1285	10	ADB76750	Adb76750 Human PRO		
c 143	17.6	70.4	1285	10	ADC44176	Adc44176 Human EST		RESULT 2
c 144	17.6	70.4	1285	10	ADC61936	Adc61936 Human EST	ADD48813/c	ADD48813/c
c 145	17.6	70.4	1285	10	ADC63900	Adc63900 Human EST	ID	ADD48813 standard; DNA; 1650 BP.
c 146	17.6	70.4	1285	10	ADC67000	Adc67000 Human EST	XX	AC
c 147	17.6	70.4	1285	10	ADC69124	Adc69124 Human EST	XX	ADD48813;
c 148	17.6	70.4	1285	10	ADC63184	Adc63184 Human EST	XX	02-DEC-2004 (revised)
c 149	17.6	70.4	1285	10	ADC68249	Adc68249 Human EST	DT	29-JAN-2004 (first entry)
c 150	17.6	70.4	1285	10	ADC41569	Adc41569 Human EST	XX	Human gene NM_016337, SEQ ID NO 14523.
							DE	Human; ds; gene; pain; neuronal tissue; gene therapy;
							KW	spinal segmental nerve injury; chronic constriction injury; CCI;
							KW	spared nerve injury; SNI; Chung.
							XX	
							OS	Homo sapiens.
							OS	Unidentified.
							XX	
							PN	WO2003016475-A2.
							XX	
							PD	27-FEB-2003.

ALIGNMENTS

RESULT 1  
 ID ABD33506 standard; DNA; 36599 BP.  
 XX  
 AC ABD33506;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Murine cancer-associated (CA) gene MD07-099.

XX 14-AUG-2002; 2002WO-US025765.  
XX 14-AUG-2001; 2001US-0312147P.  
XX 01-NOV-2001; 2001US-0346382P.  
XX 26-NOV-2001; 2001US-0333347P.  
XX (GEHO ) GEN HOSPITAL CORP.  
XX (FARB ) BAYER AG.  
XX Woelf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
XX GENBANK; NM\_016337.  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX Example 1; Page; 1017pp; English.  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a human DNA (described in Table 3 of  
XX the Specification) which encodes one of the polypeptides of the invention  
XX which is differentially expressed during pain. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1650 BP; 418 A; 488 C; 443 G; 301 T; 0 U; 0 Other;  
Query Match 80.8%; Score 20.2; DB 10; Length 1650;  
Best Local Similarity 88.0%; Pred. No. 15;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25  
Db 1393 AAAAAAGCGCAGACTGGTCTTCC 1369  
RESULT 3  
AD213146  
ID AD213146 standard; DNA; 68842 BP.  
XX AC AD213146;  
XX DT 16-JUN-2005 (first entry)  
XX DE Murine cancer-associated genomic DNA #57.  
XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;  
XX KW cytostatic; gene; ds.  
XX

OS Mus sp.  
XX WO2005031001-A2.  
XX PD 07-APR-2005.  
XX PF 23-SEP-2004; 2004WO-US031617.  
XX PR 23-SEP-2003; 2003US-00669920.  
XX (CHIR ) CHIRON CORP.  
XX Morris DW, Malandro MS;  
XX WPI; 2005-273395/28.  
XX Nucleic acid array useful for detecting cancer associated nucleic acid,  
XX comprises two or more nucleic acid probes.  
XX Disclosure; SEQ ID NO 666; 198pp; English.  
XX The invention relates to a nucleic acid array for detecting a cancer  
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
XX The invention also relates to a peptide array comprising two or more  
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
XX that binds to a polypeptide, an isolated antibody or its fragment which  
XX binds to a polypeptide, which is prepared by immunizing a host animal  
XX with a composition comprising the polypeptide or its antigen binding  
XX fragment and collecting cells from the host expressing antibodies against  
XX the antigen or its antigen binding fragment, a composition comprising the  
XX antibody and a carrier, a method of screening for anticancer activity, a  
XX method of detecting a CA nucleic acid, a method of inhibiting expression of a CA  
XX method of treating cancer and a method of inhibiting expression of a CA  
XX nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
XX nucleic acids. The antibody is useful for detecting the presence or  
XX absence of cancer cells in an individual and detecting a complex of a CA  
XX protein from the cancer cells and the antibody, where the detection of  
XX the complex correlates with the presence of cancer cells in the  
XX individual. The composition is useful for inhibiting growth of cancer  
XX cells in an individual or for delivering a therapeutic agent to cancer  
XX cells in an individual. The invention is also useful for diagnosing  
XX cancer, for treating cancer and for inhibiting expression of a CA gene in  
XX a cell. This sequence represents murine cancer-associated genomic DNA of  
XX the invention.  
XX SQ Sequence 68842 BP; 18666 A; 15495 C; 16367 G; 18314 T; 0 U; 0 Other;  
Query Match 79.2%; Score 19.8; DB 14; Length 68842;  
Best Local Similarity 91.3%; Pred. No. 49;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 AAAAAAGCGCAGACTGGTCTTCC 25  
Db 21274 AAAAAAGCGCAGACTGGTCTTCC 21296  
RESULT 4  
ACN44344  
ID ACN44344 standard; DNA; 30752 BP.  
XX AC ACN44344;  
XX AC ACN44344;  
XX DT 18-NOV-2004 (first entry)  
XX DE Mouse genomic sequence MCG17984.  
XX KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX OS Mus musculus.  
XX KW WO2003073826-A2.  
XX

PD 12-SEP-2003.  
 XX  
 PF 28-FEB-2003; 2003WO-US006235.  
 XX  
 PR 01-MAR-2002; 2002US-00087192.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 DR WPI; 2003-328604/31.  
 XX  
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 PT  
 PS Claim 1; SEQ ID NO 745; Opp; English.  
 XX  
 XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182386A1, for which no sequence data was published  
 XX  
 XX Sequence 30752 BP; 8068 A; 6394 C; 7002 G; 7660 T; 0 U; 1628 Other;  
 SQ  
  
 Query Match 77.6%; Score 19.4; DB 11; Length 30752;  
 Best Local Similarity 95.2%; Pred. No. 66;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 5 AAGAGCGCAGACTGCTCTTCC 25  
 Db 3215 AAGAGCGCAGACTGCTCTTCC 3235  
  
 RESULT 5  
 ADC85509  
 ID ADC85509 standard; DNA; 90091 BP.  
 XX  
 AC ADC85509;  
 XX  
 DT 01-JAN-2004 (first entry)  
 XX  
 DE Mouse MCG10516 genomic sequence.  
 XX  
 KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
 KW secreted; transmembrane; intracellular; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN WO2003045230-A2.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 02-DEC-2002; 2002WO-US039582.  
 XX  
 PR 30-NOV-2001; 2001US-00997722.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY.  
 PA  
 XX Morris DW, Engelhard EK;  
 PI  
 XX WPI; 2003-513603/48.  
 DR  
 XX New recombinant nucleic acid comprising a nucleotide sequence of any of

PT the carcinoma-associated (CA) genes, useful for screening for drug  
 PT candidates for diagnosing or treating carcinomas.  
 XX  
 PS Claim 1; SEQ ID NO 295; 983pp; English.  
 XX  
 XX The invention relates to a recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the fully defined carcinoma-  
 CC associated (CA) genes from the 50 tables given in the specification. The  
 CC CA proteins are secreted, transmembrane or intracellular proteins. The  
 CC recombinant nucleic acids are useful for screening for drug candidates  
 CC for diagnosing or treating carcinomas. Sequences given in ADC85215-  
 CC ADC85514 represent CA genes of the invention.  
 XX  
 XX Sequence 90091 BP; 22131 A; 13818 C; 15427 G; 24918 T; 0 U; 13797 Other;  
 SQ  
  
 Query Match 77.6%; Score 19.4; DB 10; Length 90091;  
 Best Local Similarity 95.2%; Pred. No. 81;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 5 AAGAGCGCAGACTGCTCTTCC 25  
 Db 50171 AAGAGCGCAGACTGCTCTTCC 50191  
  
 RESULT 6  
 ADA03029  
 ID ADA03029 standard; DNA; 90100 BP.  
 XX  
 AC ADA03029;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Mouse MCG10516 carcinoma associated gene, SEQ ID NO:1547.  
 XX  
 KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN WO2003057146-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 26-DEC-2002; 2002WO-US041414.  
 XX  
 PR 26-DEC-2001; 2001US-00035832.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY.  
 PA  
 XX Morris DW;  
 PI  
 XX WPI; 2003-587068/55.  
 DR  
 XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 1547; 245pp; English.  
 XX  
 XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence

CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed murine CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 90100 BP; 22131 A; 13818 C; 15428 G; 24917 T; 0 U; 13806 Other;

Query Match 77.6%; Score 19.4; DB 9; Length 90100;  
Best Local Similarity 95.2%; Pred. NO. 81;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25

DB 50180 AAGAGCGCAGACTGCTCTTCC 50200

RESULT 7

ADB72767

ID ADB72767 standard; DNA; 90100 BP.

XX AC

XX ADB72767;

XX DT 04-DEC-2003 (first entry)

XX XX

DE Mouse MCG10516 gene.

XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX OS

XX Mus sp.

XX PN WO2003008583-A2.

XX PD 30-JAN-2003.

XX PF 26-DEC-2001; 2001WO-US051291.  
XX PR 02-MAR-2001; 2001US-00798586.

XX PR 23-OCT-2001; 2001US-00004113.

XX PR 08-NOV-2001; 2001US-00052482.

XX PR 30-NOV-2001; 2001US-00997722.

XX PR 20-DEC-2001; 2001US-00034650.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX DR WPI; 2003-239337/23.

XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX PS Claim 1; SEQ ID NO 595; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined  
CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
CC sarcomas. The present sequence represents a mouse gene of the invention.

SQ Sequence 90100 BP; 22131 A; 13818 C; 15428 G; 24917 T; 0 U; 13806 Other;

Query Match 77.6%; Score 19.4; DB 10; Length 90100;

Best Local Similarity 95.2%; Pred. NO. 81;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25

DB 50180 AAGAGCGCAGACTGCTCTTCC 50200

RESULT 8

ADM74624

ID ADM74624 standard; DNA; 90100 BP.

XX AC

XX ADM74624;

XX DT 01-JUL-2004 (first entry)

XX DE Murine carcinoma associated (CA) nucleic acid #148.

XX Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;  
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;  
KW cytostatic.

XX OS

XX Mus musculus.

XX PN US2004072154-A1.

XX PD 15-APR-2004.

XX PF 30-NOV-2001; 2001US-00997722.

XX PR 22-DEC-2000; 2000US-00747377.

XX PR 02-MAR-2001; 2001US-00798586.

XX PA (MORRIS) MORRIS D W.

XX PA (ENGELHARD) ENGELHARD E K.

XX PI Morris DW, Engelhard EK;

XX DR WPI; 2004-328562/30.

XX PT New carcinoma associated gene or protein, useful for preparing a  
PT composition for diagnosing or treating carcinoma e.g., leukemia or  
PT lymphoma.

XX PS Claim 1; SEQ ID NO 295; 29pp; English.

XX The invention relates to new recombinant nucleic acids. The invention  
CC also relates to a host cell comprising a recombinant nucleic acid or  
CC expression vector, an expression vector comprising a recombinant nucleic  
CC acid, a recombinant protein, a method of screening for drug candidates, a  
CC method of screening for a bioactive agent capable of binding to a  
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
CC method of screening for a bioactive agent capable of modulating the  
CC activity of a CAP, a method of evaluating the effect of a candidate  
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting  
CC the activity of a CAP, a method of treating carcinomas, a method of  
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
CC propensity to carcinoma. A method of evaluating the effect of a candidate  
CC carcinoma drug comprises administering the drug to a patient, removing a  
CC cell sample from the patient and determining alterations in the  
CC expression or activation of a gene comprising the nucleotide sequence. A  
CC method of diagnosing carcinoma comprises determining the expression of  
CC one or more genes comprising the nucleic acid sequence in a first tissue  
CC type of a first individual and comparing the expression of the gene from  
CC a second normal tissue type from the first individual or a second  
CC unaffected individual, where a difference in the expression indicates  
CC that the first individual has carcinoma. A method of inhibiting the  
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
CC carcinomas comprises administering to a patient an inhibitor of CAP.  
CC Neutralising the effect of a CAP comprises contacting an agent specific  
CC for the CAP. The polypeptide specifically binds to the protein encoded by  
CC the nucleic acid. It comprises an antibody that specifically binds to the  
CC protein encoded by the nucleic acid. The nucleic acids are useful for  
CC preparing a composition for diagnosing or treating carcinoma e.g.,  
CC leukaemia or lymphoma. This sequence represents a murine carcinoma  
CC associated (CA) nucleic acid of the invention. Note: The sequence data  
CC for this patent did not form part of the printed specification but was

```
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 90100 BP; 22131 A; 13818 C; 15428 G; 24917 T; 0 U; 13806 Other;

Query Match 77.6%; Score 19.4; DB 12; Length 90100;
Best Local Similarity 95.2%; Pred. No. 81;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25
Db 50180 AAGAGCGCAGACTGCTCTTCC 50200

RESULT 9
AAD5112/c
ID AAD5112 standard; DNA; 1276 BP.
XX
AC AAD5112;
XX
XX 07-AUG-2003 (first entry)
XX Murine lck promoter DNA fragment.
XX Transgenic; biotechnology; agriculture; murine; promoter; ds.
XX Mus sp.
XX WO2003022040-A2.
XX
XX 20-MAR-2003.
XX
XX 12-SEP-2002; 2002WO-US029130.
XX
XX 13-SEP-2001; 2001US-0322031P.
XX 09-JAN-2002; 2002US-0347782P.
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX Baltimore D, Hong EJ, Lois-Caballe C, Pease S;
XX WPI; 2003-300976/29.
XX Producing a transgenic animal for commercial use, comprises transfecting
XX a packaging cell line with retroviral construct, recovering recombinant
XX retrovirus from the cell line and infecting embryonic cell with the
XX recombinant virus.
XX Example 1; Fig 19B; 76pp; English.
XX
XX The invention relates to a method for producing a transgenic animal for
XX commercial use, which comprises transfecting a packaging cell line with
XX retroviral construct, recovering recombinant retrovirus from the cell
XX line and infecting embryonic cell with the recombinant virus. The method
XX is useful in producing transgenic animals using retroviral constructs
XX engineered to carry a transgene of interest. The transgenic animals may
XX find use in commercial applications like biotechnology and agriculture.
XX The present sequence is murine lck promoter DNA fragment used to
XX illustrate the method of the invention
XX
XX Sequence 1276 BP; 302 A; 259 C; 381 G; 334 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 8; Length 1276;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGCAGACTGCTCTTCC 25
Db 66 AAAAAGAGCGCAGACTGCTCTTCC 43

RESULT 10
AAC45096/c
ID AAC45096 standard; DNA; 1276 BP.
XX
AC AAC45096;
XX
XX 10-JUN-2003 (first entry)
XX Murine lck promoter partial nucleotide sequence SEQ ID NO:6.
XX Transgenic animal; transgenic bird; transgenic fish; transgene;
XX retroviral construct; lentiviral; long terminal repeat; LTR;
XX biotechnology; agriculture; murine; lck promoter; ds.
XX Mus sp.
XX WO2003022228-A2.
XX
XX 20-MAR-2003.
XX
XX 12-SEP-2002; 2002WO-US029157.
XX
XX 13-SEP-2001; 2001US-0322031P.
XX 09-JAN-2002; 2002US-0347782P.
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX Baltimore D, Hong EJ, Lois-Caballe C, Pease S;
XX WPI; 2003-301005/29.
XX Producing a transgenic bird or fish for commercial use, comprises
XX transfecting a packaging cell line with retroviral construct, recovering
XX recombinant retrovirus from the cell and infecting bird or fish egg with
XX the recombinant virus.
XX Example 1; Fig 19; 68pp; English.
XX
XX The present invention describes a method for producing a transgenic bird
XX or fish. The method comprises transfecting a packaging cell line with a
XX retroviral construct, recovering recombinant retroviral particles from
XX the packaging cell line, and infecting a bird or a fish egg with the
XX recombinant retroviral particles. The retroviral construct comprises the
XX R and U5 sequences from a 5' lentiviral long terminal repeat (LTR) and a
XX self-inactivating 3' lentiviral LTR. Also described is a transgenic bird
XX or fish made by the above method and whose genome comprises a proviral
XX DNA that has a self-inactivating 3' lentiviral LTR. The method is useful
XX in producing transgenic animals, particularly transgenic birds and fish,
XX using retroviral constructs engineered to carry a transgene of interest.
XX The method is used to introduce the gene of choice into animals in order
XX to confer upon them desired attributes. The transgenic animals may find
XX use in commercial applications like biotechnology and agriculture. The
XX present sequence represents a murine lck promoter partial nucleotide
XX sequence, which is used in an example from the present invention
XX
XX Sequence 1276 BP; 302 A; 259 C; 381 G; 334 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 8; Length 1276;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGCAGACTGCTCTTCC 25
Db 66 AAAAAGAGCGCAGACTGCTCTTCC 43

RESULT 11
ACA36353/c
ID ACA36353 standard; DNA; 1392 BP.
XX
AC ACA36353;
XX
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #18010.
XX
```

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX Listeria monocytogenes.  
 OS WO200277183-A2.  
 PN 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US0009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JB, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR P-PSDB; ABU32483.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 14; SEQ ID NO 24223; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1392 BP; 442 A; 230 C; 346 G; 374 T; 0 U; 0 Other;  
 SQ Query Match 76.8%; Score 19.2; DB 8; Length 1392;  
 Best Local Similarity 87.5%; Pred. No. 46;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AAAAAAGAGCGCAGACTGGTCTTC 24  
 |||||||  
 Db 132 AAAAAAGAGCGCAGACTGGTCTTC 109  
 |||||||

RESULT 12  
 ADA02963/c  
 ID ADA02963 standard; DNA; 29956 BP.  
 XX AC ADA02963;  
 XX 06-NOV-2003 (first entry)  
 DT Mouse Lck carcinoma associated gene, SEQ ID NO:1481.  
 DE Mouse; murine; carcinoma associated; oncogene; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ds.  
 XX OS Mus sp.  
 XX WO2003057146-A2.  
 FN 17-JUL-2003.  
 PD 26-DEC-2002; 2002WO-US041414.  
 PF 26-DEC-2001; 2001US-00035832.  
 PR (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW;  
 PI WPI; 2003-587068/55.  
 DR New recombinant nucleic acid encoding carcinoma associated protein,  
 XX useful for preparing compositions for treating carcinomas.  
 PT Claim 1; SEQ ID NO 1481; 245pp; English.  
 PS The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 29956 BP; 6634 A; 6708 C; 6677 G; 7953 T; 0 U; 1984 Other;  
 SQ Query Match 76.8%; Score 19.2; DB 9; Length 29956;  
 Best Local Similarity 87.5%; Pred. No. 82;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 AAAAAAGAGCGCAGACTGGTCTTC 25  
 |||||||  
 Db 8863 AAAAAAGAGCGCAGACTGGTCTTC 8840  
 |||||||

RESULT 13  
 ADB72701/c



```

ID ADB72701 standard; DNA; 29956 BP.
XX
AC ADB72701;
XX
DT 04-DEC-2003 (first entry)
XX
DE Mouse Lck gene.
XX
KW mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Mus sp.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 529; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ Sequence 29956 BP; 6634 A; 6708 C; 6677 G; 7953 T; 0 U; 1984 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 29956;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGAGCTGCTCTTCC 25
DB 8863 AAAAAGAGCACTGACTGCTCTTCC 8840

RESULT 14
ADC85443/c
ID ADC85443 standard; DNA; 29956 BP.
XX
AC ADC85443;
XX
DT 01-JAN-2004 (first entry)
XX
DE Mouse Lck genomic sequence.
XX
KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
OS Mus sp.
XX
PN WO2003045230-A2.
XX
PD 05-JUN-2003.

Query Match 76.8%; Score 19.2; DB 10; Length 29956;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGAGCTGCTCTTCC 25
DB 8863 AAAAAGAGCACTGACTGCTCTTCC 8840

RESULT 15
ADM74558/c
ID ADM74558 standard; DNA; 29956 BP.
XX
AC ADM74558;
XX
DT 01-JUL-2004 (first entry)
XX
DE Murine carcinoma associated (CA) nucleic acid #115.
XX
KW Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW cytostatic.
XX
OS Mus musculus.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
PWPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 229; 299pp; English.
```

XX The invention relates to new recombinant nucleic acids. The invention  
CC also relates to a host cell comprising a recombinant nucleic acid or  
CC expression vector, an expression vector comprising a recombinant nucleic  
CC acid, a recombinant protein, a method of screening for drug candidates, a  
CC method of screening for a bioactive agent capable of binding to a  
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
CC method of screening for a bioactive agent capable of modulating the  
CC activity of a CAP, a method of evaluating the effect of a candidate  
CC carcinoma drug, a method of diagnosing carcinomas, a method for inhibiting  
CC the activity of a CAP, a method of treating carcinomas, a method of  
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
CC propensity to carcinoma. A method of evaluating the effect of a candidate  
CC carcinoma drug comprises administering the drug to a patient, removing a  
CC cell sample from the patient and determining alterations in the  
CC expression or activation of a gene comprising the nucleotide sequence. A  
CC method of diagnosing carcinoma comprises determining the expression of  
CC one or more genes comprising the nucleic acid sequence in a first tissue  
CC type of a first individual and comparing the expression of the gene from  
CC a second normal tissue type from the first individual or a second  
CC unaffected individual, where a difference in the expression indicates  
CC that the first individual has carcinoma. A method of inhibiting the  
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
CC carcinomas comprises administering to a patient an inhibitor of CAP.  
CC Neutralising the effect of a CAP comprises contacting an agent specific  
CC for the CAP. The polypeptide specifically binds to the protein encoded by  
CC the nucleic acid. It comprises an antibody that specifically binds to the  
CC protein encoded by the nucleic acid. The nucleic acids are useful for  
CC preparing a composition for diagnosing or treating carcinoma e.g.,  
CC leukemia or lymphoma. This sequence represents a murine carcinoma  
CC associated (CA) nucleic acid of the invention. Note: The sequence data  
CC for this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

XX SQ Sequence 29956 BP; 6634 A; 6708 C; 6677 G; 7953 T; 0 U; 1984 Other;

Query Match 76.8%; Score 19.2; DB 12; Length 29956;  
Best Local Similarity 87.5%; Pred. No. 82;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAGAGCGCAGACTGGTCTTC 25  
Db 8863 AAAAAAGAGCGCAGCTGCTCTTC 8840

RESULT 16

ABA03041\_24 Continuation (25 of 30) of ABA03041 from base 2400001 (Listeria monocytogenes EGD-e gene

WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

WP Fragment Name Begin End

WP ABA03041\_00 1 110000

WP ABA03041\_01 100001 210000

WP ABA03041\_02 200001 310000

WP ABA03041\_03 300001 410000

WP ABA03041\_04 400001 510000

WP ABA03041\_05 500001 610000

WP ABA03041\_06 600001 710000

WP ABA03041\_07 700001 810000

WP ABA03041\_08 800001 910000

WP ABA03041\_09 900001 1010000

WP ABA03041\_10 1000001 1110000

WP ABA03041\_11 1100001 1210000

WP ABA03041\_12 1200001 1310000

WP ABA03041\_13 1300001 1410000

WP ABA03041\_14 1400001 1510000

WP ABA03041\_15 1500001 1610000

WP ABA03041\_16 1600001 1710000

WP ABA03041\_17 1700001 1810000

WP ABA03041\_18 1800001 1910000

WP ABA03041\_19 1900001 2010000

WP ABA03041\_20 2000001 2110000

WP ABA03041\_21 2100001 2210000

WP ABA03041\_22 2200001 2310000  
WP ABA03041\_23 2300001 2410000  
WP ABA03041\_24 2400001 2510000  
WP ABA03041\_25 2500001 2610000  
WP ABA03041\_26 2600001 2710000  
WP ABA03041\_27 2700001 2810000  
WP ABA03041\_28 2800001 2910000  
WP ABA03041\_29 2900001 2944528

Query Match 76.8%; Score 19.2; DB 6; Length 110000;

Best Local Similarity 87.5%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTC 24

Db 77120 AAAAAAGAGCGCGACTGATTTTC 77143

RESULT 17

ADV34995

ID ADV34995 standard; cDNA; 123192 BP.

XX AC ADV34995;

XX DT 10-FEB-2005 (first entry)

XX DE Murine cDNA differentially expressed in the presence of valproate Seq71.  
XX KW murine; mouse; valproate; ss; multi-parameter high throughput screening;  
XX KW MPHS; disease signature; neuropsychiatric; neurodegenerative;  
XX KW schizophrenia; bipolar affective disorder; BAD; autism; Parkinson's;  
XX KW Alzheimer's disease; neuroleptic; nootropic; antimanic; antidepressa.

XX OS Mus musculus.

XX PN US2003096264-A1.

XX PD 22-MAY-2003.

XX PF 18-JUN-2002; 2002US-00175523.

XX PR 18-JUN-2001; 2001US-0299151P.

XX PR 07-SEP-2001; 2001US-0317828P.

XX PR 25-SEP-2001; 2001US-0325150P.

XX PR 14-NOV-2001; 2001US-0333047P.

XX PR 18-JAN-2002; 2002US-0349936P.

XX PR 04-MAR-2002; 2002US-0361834P.

XX (PSYC-) PSYCHIATRIC GENOMICS INC.

XX Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Laeng P;

XX Palfreyman M, Rajan P;

XX WPI; 2004-118903/12.

XX Identifying a compound that can treat disease or disorders, such as, a

XX neuropsychiatric disorder e.g. schizophrenia, or autism, comprises

XX determining the expression of one or more efficacy genes in a cell

XX contacted with the test compound.

XX Claim 9; SEQ ID NO 71; 39pp; English.

XX This invention relates to a novel screening method identified as a multi-  
XX parameter high throughput screening (MPHS) assay. Specifically, it  
XX refers to an assay that utilises the disease signature of a plurality of  
XX specific genes associated with a particular disease, and identifies  
XX differential expression between those cells taken from individuals  
XX affected by that disease and those that are not affected. The present  
XX invention then describes the screening of candidate pharmaceutical  
XX compounds to identify those that have a potential therapeutic benefit for  
XX the treatment of neuropsychiatric and neurodegenerative disorders  
XX including schizophrenia, bipolar affective disorder (BAD) and autism, as  
XX well as Parkinson's and Alzheimer's disease. Accordingly, the compounds

CC of this invention exhibit various activities including neuroleptic,  
CC nootropic, antimanic and antidepressant. Furthermore, the screening  
CC method used in MPHTS will be automated, such that a large number of test  
CC compounds may be rapidly screened with a minimal amount of labour and  
CC effort. This polynucleotide is the cDNA sequence of a gene that is  
CC differentially expressed in mice in the presence of the therapeutic  
CC compound valproate, given in an exemplification of the invention.

XX  
SQ Sequence 123192 BP; 37038 A; 25773 C; 25891 G; 34490 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 13; Length 123192;  
Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGCTTC 24  
Db 73891 AAAAAAGAGCACTGACTGCTTC 73914  
|||||

RESULT 18  
ADH18345  
ID ADH18345 standard; DNA; 43445 BP.  
AC ADH18345;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Human apolipoprotein B (ApoB) DNA 3 - SEQ ID 334.

XX apolipoprotein B; ApoB; antiarteriosclerotic; cardiant; antidiabetic;  
XX anorectic; lipid; cholesterol metabolism; atherosclerosis;  
KW diabetes Type 2; obesity; hyperlipidaemia; cardiovascular; gene therapy;  
KW human; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2003097662-A1.  
XX  
XX 27-NOV-2003.  
XX  
XX 15-MAY-2003; 2003WO-US015493.  
XX  
XX 15-MAY-2002; 2002US-00147196.  
PR 13-NOV-2002; 2002US-0426224P.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Crooke RM, Graham MJ;  
XX  
XX WPI; 2004-022840/02.

XX New antisense compound, useful for preparing a composition for treating  
PT abnormal lipid or cholesterol metabolism, atherosclerosis, diabetes Type  
PT 2, obesity, hyperlipidemia or cardiovascular disease.  
XX  
XX Example 33; SEQ ID NO 334; 405pp; English.

XX  
XX The invention relates to a novel antisense compound targeted to a nucleic  
CC acid molecule encoding human apolipoprotein B (ApoB) which specifically  
CC hybridises with and inhibits the expression of human apolipoprotein B.  
CC The compound of the invention demonstrates antiarteriosclerotic,  
CC cardiant, antidiabetic and anorectic activities and may be useful for  
CC preparing a composition for treating abnormal lipid or cholesterol  
CC metabolism, atherosclerosis, diabetes Type 2, obesity, hyperlipidaemia or  
CC cardiovascular disease. Furthermore, the compound has gene therapy  
CC applications. The current sequence is that of the human ApoB DNA of the  
CC invention.

XX  
SQ Sequence 43445 BP; 12661 A; 9313 C; 9323 G; 12148 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 12; Length 43445;  
Best Local Similarity 90.9%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGCTTC 22  
Db 25768 AAAAAAGAGCGTAGACTGCTCT 25789  
|||||

RESULT 19  
ADO32886  
ID ADO32886 standard; DNA; 43445 BP.  
XX  
AC ADO32886;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE Human apolipoprotein B (ApoB) gDNA - SEQ ID 334.

XX apolipoprotein B; ApoB; cardiovascular; antiarteriosclerotic;  
KW antilipaeamic; antidiabetic; anorectic; cardiant; vasotropic; hypotensive;  
KW anabolic; eating disorder; cytostatic; endocrine; vasotropic;  
KW neuroprotective; nootropic; lipid; cholesterol metabolism;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW Von Gierke's disease; lipodystrophy; Cushing's syndrome;  
KW sexual ateliotic dwarfism; hyperthyroidism; hypertension;  
KW anorexia nervosa; Werner's syndrome; hepatoma; multiple myeloma; uraemia;  
KW impotence; obstructive liver disease; Alzheimer's; dementia; diabetes;  
KW obesity; atherosclerosis; human; ds; chromosome 2p23-2p24.

XX  
XX Homo sapiens.  
XX  
XX WO2004044181-A2.  
XX  
XX 27-MAY-2004.  
XX  
XX 13-NOV-2003; 2003WO-US036411.  
XX  
XX 13-NOV-2002; 2002US-0426234P.  
PR 15-MAY-2003; 2003WO-US015493.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Crooke R, Graham M, Lemonidis-Tarbet K, Dobie KW;  
XX  
XX WPI; 2004-420321/39.

XX Antisense oligonucleotide compound that inhibits expression of mRNA  
PT encoding human apolipoprotein B, useful for treating hyperlipidemia,  
PT diabetes, obesity, von Gierke's disease, lipodystrophies, Cushing's  
PT syndrome.  
XX  
XX Example 33; SEQ ID NO 334; 483pp; English.

XX  
XX The invention relates to a novel antisense compound where the compound  
CC hybridises to and inhibits expression of mRNA encoding human  
CC apolipoprotein B (ApoB) after 16-24 hours by at least 30% in 80%  
CC confluent HepG2 cells in culture at a concentration of 150 nM. The  
CC compound of the invention demonstrates cardiovascular,  
CC antiarteriosclerotic, antilipaeamic, antidiabetic, anorectic, cardiant,  
CC vasotropic, hypotensive, anabolic, eating disorder-related, cytostatic,  
CC endocrine, vasotropic, neuroprotective and nootropic activities and may  
CC be useful for inhibiting the expression of apolipoprotein B in cells or  
CC tissues in vivo in order to address a condition associated with abnormal  
CC lipid or cholesterol metabolism. The compound may be useful for  
CC decreasing circulating lipoprotein levels, triglyceride levels,  
CC cholesterol levels, lipid levels, fatty acid levels, acute phase  
CC reactants and chylomicrons and thus may be utilised during treatment of  
CC hyperlipoproteinaemia, hyperlipidaemia, hypercholesterolaemia,  
CC cardiovascular disorders, Von Gierke's disease, lipodystrophy, Cushing's  
CC syndrome, sexual ateliotic dwarfism, hyperthyroidism, hypertension,  
CC anorexia nervosa, Werner's syndrome, hepatoma, multiple myeloma, uraemia,  
CC impotence, obstructive liver disease, Alzheimer's disease, dementia, of  
CC diabetes, obesity and atherosclerosis. The current sequence is that of  
CC the human apolipoprotein B (ApoB) gDNA - SEQ ID 334 of the invention  
CC which is located at chromosome 2p23-2p24.

```
XX SQ Sequence 43445 BP; 12661 A; 9314 C; 9322 G; 12148 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 12; Length 43445;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCT 22
||||| ||||| ||||| ||||| |||||
Db 25768 AAAAAAGCGCTAGACTGGTCT 25789

RESULT 20
ACN44704
ID ACN44704 standard; DNA; 39405 BP.
XX AC ACN44704;
XX
DT 18-NOV-2004 (first entry)
XX
DE Mouse genomic sequence MCG2509.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX
OS Mus musculus.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 1285; Opp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for treating
CC carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip;
CC carcinoma; (viii) for neutralizing the effect of CAP; (x) for
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 39405 BP; 10420 A; 8422 C; 8690 G; 11580 T; 0 U; 293 Other;
Query Match 74.4%; Score 18.6; DB 11; Length 39405;
Best Local Similarity 84.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 25
||||| ||||| ||||| ||||| |||||
Db 9488 AAAAAAGAGCACTGCTCTTC 9512

RESULT 21
ACN44796
```

```
ID ACN44796 standard; DNA; 39790 BP.
XX AC ACN44796;
XX
DT 18-NOV-2004 (first entry)
XX
DE Mouse genomic sequence MCG4192.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX
OS Mus musculus.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 1423; Opp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for treating
CC carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip;
CC carcinoma; (viii) for neutralizing the effect of CAP; (x) for
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 39790 BP; 9126 A; 9461 C; 9319 G; 8925 T; 0 U; 2959 Other;
Query Match 74.4%; Score 18.6; DB 11; Length 39790;
Best Local Similarity 84.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 25
||||| ||||| ||||| ||||| |||||
Db 15595 AATAAAGAGCACTGCTCTTC 15619

RESULT 22
ADA02627/c
ID ADA02627 standard; DNA; 47115 BP.
XX AC ADA02627;
XX
DT 06-NOV-2003 (first entry)
XX
DE Mouse Flt3 carcinoma associated gene, SEQ ID NO:1145.
XX
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Mus sp.
```

PN WO2003057146-A2.  
 XX 17-JUL-2003.  
 PD 26-DEC-2002; 2002WO-US041414.  
 XX 26-DEC-2001; 2001US-00035832.  
 PF (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW;  
 XX WPI; 2003-587068/55.  
 DR New recombinant nucleic acid encoding carcinoma associated protein,  
 XX useful for preparing compositions for treating carcinomas.  
 PT Claim 1; SEQ ID NO 1145; 245pp; English.  
 XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 47115 BP; 11483 A; 10181 C; 10540 G; 13012 T; 0 U; 1899 Other;  
 Query Match 74.4%; Score 18.6; DB 9; Length 47115;  
 Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAAGAGCGAGACTGGTCTTCC 25  
 Db 5679 AAAAAAGAGCATGGACTGCTCTTCC 5655  
 RESULT 23  
 ADB72365/c  
 ID ADB72365 standard; DNA; 47115 BP.  
 XX ADB72365;  
 AC 04-DEC-2003 (first entry)  
 XX Mouse Flt3 gene.  
 DE mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
 KW Mus sp.  
 OS WO2003008583-A2.  
 XX 30-JAN-2003.  
 PD 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.  
 PR 23-OCT-2001; 2001US-00004113.  
 PR 08-NOV-2001; 2001US-00052482.  
 PR 30-NOV-2001; 2001US-00997722.  
 PR 20-DEC-2001; 2001US-00034650.  
 XX (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW, Engelhard EK;  
 PI WPI; 2003-239337/23.  
 DR New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 XX cancers, neoplasm, adenocarcinoma, or sarcomas.  
 PT Claim 1; SEQ ID NO 193; 2304pp; English.  
 XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a mouse gene of the invention.  
 XX SQ Sequence 47115 BP; 11483 A; 10181 C; 10540 G; 13012 T; 0 U; 1899 Other;  
 Query Match 74.4%; Score 18.6; DB 10; Length 47115;  
 Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 AAAAAAGAGCGAGACTGGTCTTCC 25  
 Db 5679 AAAAAAGAGCATGGACTGCTCTTCC 5655  
 RESULT 24  
 ADE95875/c  
 ID ADE95875 standard; DNA; 47115 BP.  
 XX ADE95875;  
 AC 12-FEB-2004 (first entry)  
 XX Mouse Flt3 gene genomic DNA sequence.  
 DE cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;  
 KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; mouse; murine;  
 KW Flt3.  
 XX Mus sp.  
 OS WO2003039484-A2.  
 XX 15-MAY-2003.  
 PD 08-NOV-2002; 2002WO-US036071.  
 PF 08-NOV-2001; 2001US-00052482.  
 XX (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW, Engelhard EK;  
 PI WPI; 2003-441462/41.  
 DR New carcinoma associated nucleic acids and proteins, useful for screening  
 XX drug candidates, or for diagnosing and treating carcinomas, e.g.  
 PT lymphoma, breast cancer, prostate cancer or leukemia.  
 XX Claim 1; SEQ ID NO 133; 793pp; English.

CC This invention relates to novel recombinant nucleic acids for use in  
CC diagnosis and treatment of cancer, especially carcinomas, as well as the  
CC use of compositions in screening methods. The compositions of the  
CC invention may have cytostatic activity whilst the disclosed sequences may  
CC be useful for gene therapy. The carcinoma associated nucleic acids and  
CC proteins are useful for diagnosing and treating carcinomas, for example  
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
CC drug candidates or bioactive agents capable of binding to, or modulating  
CC the activity of, a carcinoma associated protein. The present sequence is  
CC the genomic DNA sequence of the mouse Flt3 gene which is a carcinoma  
CC associated gene of the invention.

XX SQ Sequence 47115 BP; 11483 A; 10181 C; 10539 G; 13012 T; 0 U; 1900 Other;

Query Match 74.4%; Score 18.6; DB 10; Length 47115;  
Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGCTCTTCC 25  
DB 5679 AAAAAAGAGCGACACTGCTCTTCC 5655

RESULT 25  
ACN44048/c  
ID ACN44048 standard; DNA; 49281 BP.

XX AC ACN44048;

XX XX 18-NOV-2004 (first entry)

XX DE Mouse genomic sequence MCG13312.

XX KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX OS Mus musculus.

XX PN W02003073826-A2.

XX PD 12-SEP-2003.

XX PF 28-FEB-2003; 2003WO-US006235.

XX PR 01-MAR-2002; 2002US-00087192.

XX XX (SAGR-) SAGRES DISCOVERY.

XX PA Morris DW;

XX PI WPI; 2003-328604/31.

XX DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX PT comprises a nucleotide sequence.

XX PS Claim 1; SEQ ID NO 301; Opp; English.

XX CC The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published

XX SQ Sequence 49281 BP; 12534 A; 8091 C; 8876 G; 13885 T; 0 U; 5895 Other;

Query Match 74.4%; Score 18.6; DB 11; Length 49281;  
Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGCTCTTCC 25  
DB 11432 AAAAAAGAGCGACACTGCTCTTCC 11408

RESULT 26

ADQ59515/c

ID ADQ59515 standard; DNA; 64482 BP.

XX AC ADQ59515;

XX DT 07-OCT-2004 (first entry)

XX DE Human cancer-associated (CA) gene sequence SEQ ID NO:151.

XX KW human; cancer-associated gene; cancer-associated protein; cytostatic;  
XX KW gene therapy; vaccine; tyrosine kinase antagonist;  
XX KW G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.

XX OS Homo sapiens.

XX PN W02004058288-A1.

XX PD 15-JUL-2004.

XX PF 15-DEC-2003; 2003WO-US040082.

XX PR 17-DEC-2002; 2002US-00322696.

XX XX (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Malandro MS;

XX XX WPI; 2004-543349/52.

XX DR P-PSDB; ADQ59517.

XX XX New cancer-associated nucleic acid for diagnosing, preventing or treating  
XX cancer (e.g. lymphoma) or for screening agents that may be used for  
XX treating or preventing cancer.

XX PS Disclosure; SEQ ID NO 151; 143pp; English.

XX CC The present invention describes human cancer-associated (CA) nucleotide  
XX sequences (i). Also described: (1) an expression vector comprising (1);  
XX (2) a host cell comprising (1) or the expression vector; (3) a microarray  
XX for detecting a CA nucleic acid; (4) an isolated polypeptide encoded  
XX within an open reading frame of a CA sequence; (5) an isolated antibody,  
XX or its antigen binding fragment, that binds to the above polypeptide; (6)  
XX a hybridoma that produces the monoclonal antibody described above; (7) a  
XX pharmaceutical composition comprising the antibody and a pharmaceutical  
XX excipient; (8) a kit for detecting or diagnosing cancer cells, comprising  
XX the above (monoclonal) antibody or polynucleotide that selectively  
XX hybridizes to any of the polynucleotide sequences mentioned above; (9)  
XX methods for diagnosing cancer or for detecting the presence or absence of  
XX cancer cells in an individual; (10) a method for inhibiting growth of  
XX cancer cells in an individual; (11) a method for delivering a therapeutic  
XX agent to cancer cells in an individual; (12) an electronic library  
XX comprising the polynucleotide or polypeptide, or their fragments,  
XX mentioned above; (13) a method of screening for anticancer activity; (14)  
XX methods for detecting cancer associated with expression of a polypeptide  
XX or the presence of the antibody in a test cell or serum sample; (15) a  
XX method for screening for a bioactive agent capable of modulating the  
XX activity of a CA protein encoded by the above nucleic acid molecule; and  
XX (16) a method for treating cancers. (i) has cytostatic activity, and can  
XX be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,  
XX and as a G-protein coupled receptor antagonist. The compositions and  
XX methods of the present invention can be used for diagnosing, preventing  
XX and treating cancer, especially lymphomas. They may also be used in  
XX screening for agents that may be used for treating or preventing cancer.

CC The present sequence represents a human CA gene sequence, which is given  
CC in the exemplification of the present invention. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 64482 BP; 15643 A; 14684 C; 15303 G; 17456 T; 0 U; 1396 Other;  
SQ

Query Match 74.4%; Score 18.6; DB 12; Length 64482;

Best Local Similarity 84.0%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAGAGCGAGACTGCTCTTCC 25

Db 53861 AATAAAGAGCAGTGTCTTCC 53837

RESULT 27

ADZ13896/c

ID ADZ13896 standard; DNA; 64482 BP.

XX AC ADZ13896;

XX DT 16-JUN-2005 (first entry)

XX DE Murine cancer-associated genomic DNA #123.

XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;

XX KW cytostatic; gene; ds.

XX OS Mus sp.

XX PN WO2005031001-A2.

XX PD 07-APR-2005.

XX PF 23-SEP-2004; 2004WO-US031617.

XX PR 23-SEP-2003; 2003US-00669920.

XX PA (CHIR ) CHIRON CORP.

XX PI Morris DW, Malandro MS;

XX DR WPI; 2005-273395/28.

XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,  
XX comprises two or more nucleic acid probes.  
XX Disclosure; SEQ ID NO 1416; 198pp; English.  
XX The invention relates to a nucleic acid array for detecting a cancer  
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
XX The invention also relates to a peptide array comprising two or more  
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
XX that binds to a polypeptide, an isolated antibody or its fragment which  
XX binds to a polypeptide, which is prepared by immunizing a host animal  
XX with a composition comprising the polypeptide or its antigen binding  
XX fragment and collecting cells from the host expressing antibodies against  
XX the antigen or its antigen binding fragment, a composition comprising the  
XX antibody and a carrier, a method of screening for anticancer activity, a  
XX method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
XX method of treating cancer and a method of inhibiting expression of a CA  
XX nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
XX nucleic acids. The antibody is useful for detecting the presence or  
XX absence of cancer cells in an individual which involves contacting cells  
XX from the individual with the antibody and detecting a complex of a CA  
XX protein from the cancer cells and the antibody, where the detection of  
XX the complex correlates with the presence of cancer cells in the  
XX individual. The composition is useful for inhibiting growth of cancer  
XX cells in an individual or for delivering a therapeutic agent to cancer  
XX cells in an individual. The invention is also useful for diagnosing  
XX cancer, for treating cancer and for inhibiting expression of a CA gene in

CC a cell. This sequence represents murine cancer-associated genomic DNA of  
CC the invention.

XX SQ Sequence 64482 BP; 15643 A; 14684 C; 15303 G; 17456 T; 0 U; 1396 Other;  
SQ

Query Match 74.4%; Score 18.6; DB 14; Length 64482;

Best Local Similarity 84.0%; Pred. No. 1.9e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAGAGCGAGACTGCTCTTCC 25

Db 53861 AATAAAGAGCAGTGTCTTCC 53837

RESULT 28

ABD33614/c

ID ABD33614 standard; DNA; 89213 BP.

XX AC ABD33614;

XX DT 18-NOV-2004 (first entry)

XX DE Murine cancer-associated (CA) gene MD07-125.

XX KW Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;

XX KW ds; cancer; cytostatic.

XX OS Mus musculus.

XX PN WO2004058146-A2.

XX PD 15-JUL-2004.

XX PF 15-DEC-2003; 2003WO-US040081.

XX PR 17-DEC-2002; 2002US-00322281.

XX PA (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Malandro MS;

XX DR WPI; 2004-499109/47.

XX PT Novel human cancer associated protein encoded within open reading frame

XX of cancer associated gene, useful as targets for diagnosing cancer.

XX PS Disclosure; SEQ ID NO 843; 182pp; English.

XX CC The invention relates to cancer-associated proteins (CAP) and the cancer-

XX associated (CA) nucleic acids encoding them. The invention also relates

XX to a method for treating cancers involving administering to a patient an

XX inhibitor of CAP, and a method of screening for anticancer activity in a

XX potential drug involving providing a cell that expresses a CA gene,

XX contacting a tissue sample derived from a cancer cell with an anticancer

XX drug candidate and monitoring the effect of the anticancer drug candidate

XX on expression of the CA gene. The CAP proteins are useful for detecting

XX cancer associated with expression of a CAP protein in a test cell sample

XX and for screening for a bioactive agent capable of modulating the

XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing

XX cancer, involving determining the expression of a CA nucleic acid in a

XX tissue. This sequence represents a murine CA gene of the invention. Note:

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 89213 BP; 24812 A; 17038 C; 18589 G; 27735 T; 0 U; 1039 Other;

XX

Query Match 74.4%; Score 18.6; DB 13; Length 89213;

Best Local Similarity 84.0%; Pred. No. 2e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAGAGCGAGACTGCTCTTCC 25

Db 53861 AATAAAGAGCAGTGTCTTCC 53837



```
Db      81877 AAAAAAGAGTGTCGACTGCTCTTCC 81853

RESULT 29
AD212547/c
ID      AD212547 standard; DNA; 92099 BP.
XX
AC      AD212547;
XX
DT      16-JUN-2005 (first entry)
XX
DE      Murine cancer-associated genomic DNA #8.
XX
KW      Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW      cytostatic; gene; ds.
XX
OS      Mus sp.
XX
PN      W02005031001-A2.
XX
PD      07-APR-2005.
XX
PF      23-SEP-2004; 2004WO-US031617.
XX
PR      23-SEP-2003; 2003US-00669920.
XX
PA      (CHIR ) CHIRON CORP.
XX
PI      Morris DW, Malandro MS;
XX
WPI; 2005-273395/28.
XX
Nucleic acid array useful for detecting cancer associated nucleic acid,
PT      comprises two or more nucleic acid probes.
XX
PS      Disclosure; SEQ ID NO 67; 198pp; English.
XX
CC      The invention relates to a nucleic acid array for detecting a cancer
CC      associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC      The invention also relates to a peptide array comprising two or more
CC      isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC      that binds to a polypeptide, an isolated antibody or its fragment which
CC      binds to a polypeptide, which is prepared by immunizing a host animal
CC      with a composition comprising the polypeptide or its antigen binding
CC      fragment and collecting cells from the host expressing antibodies against
CC      the antigen or its antigen binding fragment, a composition comprising the
CC      antibody and a carrier, a method of screening for anticancer activity, a
CC      method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC      method of treating cancer and a method of inhibiting expression of a CA
CC      nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC      nucleic acids. The antibody is useful for detecting the presence or
CC      absence of cancer cells in an individual which involves contacting cells
CC      from the individual with the antibody and detecting a complex of a CA
CC      protein from the cancer cells and the antibody, where the detection of
CC      the complex correlates with the presence of cancer cells in the
CC      individual. The composition is useful for inhibiting growth of cancer
CC      cells in an individual or for delivering a therapeutic agent to cancer
CC      cells in an individual. The invention is also useful for diagnosing
CC      cancer, for treating cancer and for inhibiting expression of a CA gene in
CC      a cell. This sequence represents murine cancer-associated genomic DNA of
CC      the invention.
XX
SQ      Sequence 92099 BP; 23678 A; 20842 C; 20864 G; 25593 T; 0 U; 1122 Other;

Query Match      74.4%; Score 18.6; DB 14; Length 92099;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1 AAAAAAGAGCGCAGACTGCTCTTCC 25
      |||||
Db      49960 AAAAAAGAGCATGGACTGCTCTTCC 49936

RESULT 30
ADA02627
ID      ADA02627 standard; DNA; 47115 BP.
XX
AC      ADA02627;
XX
DT      06-NOV-2003 (first entry)
XX
DE      Mouse Flt3 carcinoma associated gene, SEQ ID NO:1145.
XX
KW      Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW      prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW      gene; ds.
XX
OS      Mus sp.
XX
PN      W02003057146-A2.
XX
PD      17-JUL-2003.
XX
PF      26-DEC-2002; 2002WO-US041414.
XX
PR      26-DEC-2001; 2001US-00035832.
XX
PA      (SAGR-) SAGRES DISCOVERY.
XX
PI      Morris DW;
XX
WPI; 2003-587068/55.
XX
New recombinant nucleic acid encoding carcinoma associated protein,
PT      useful for preparing compositions for treating carcinomas.
XX
PS      Claim 1; SEQ ID NO 1145; 245pp; English.
XX
CC      The invention relates to recombinant carcinoma associated (CA) nucleic
CC      acid sequences from mouse and human (ADA01482-ADA03094), and to
CC      recombinant carcinoma associated proteins (CAP) encoded by them. The
CC      invention also encompasses expression vectors and host cells comprising a
CC      CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC      binds to the protein, and a biochip comprising CA nucleic acid or
CC      fragments thereof. The sequences of the invention were identified using
CC      oncogenic retroviruses, which insert into the genome of the host organism
CC      at random. Many of these do not carry transduced host oncogenes or
CC      pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC      direct consequence of the effects of proviral integration into host
CC      protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC      carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC      leukaemia) or a propensity to carcinoma by determination of the sequence
CC      of a CA gene, or by determination of CA gene expression in particular
CC      tissues. CA nucleic acids, proteins and antibodies are also useful as
CC      therapeutic agents and in screening and evaluating drug candidates. The
CC      present sequence represents a specifically claimed murine CA nucleic acid
CC      sequence of the invention. Note: The complete sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 47115 BP; 11483 A; 10181 C; 10540 G; 13012 T; 0 U; 1899 Other;

Query Match      73.6%; Score 18.4; DB 9; Length 47115;
Best Local Similarity 95.0%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      5 AAGAGCGCAGACTGCTCTTC 24
      |||||
Db      34115 AAGAGCGCAGACTGCTCTTC 34134

RESULT 31
ADB72365
ID      ADB72365 standard; DNA; 47115 BP.
XX
```

```
AC ADB72365;
XX
DT 04-DEC-2003 (first entry)
XX
XX DE Mouse Flt3 gene.
XX
XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Mus sp.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
XX
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW, Engelhard EK;
PI
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
PT
XX Claim 1; SEQ ID NO 193; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ Sequence 47115 BP; 11483 A; 10181 C; 10540 G; 13012 T; 0 U; 1899 Other;
Query Match 73.6%; Score 18.4; DB 10; Length 47115;
Best Local Similarity 95.0%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTC 24
Db 34115 AAGAGCGCAGACTGCTCTTC 34134

RESULT 32
ADE95875
ID ADE95875 standard; DNA; 47115 BP.
XX
AC ADE95875;
XX
XX 12-FEB-2004 (first entry)
DT
DE Mouse Flt3 gene genomic DNA sequence.
XX
XX cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; mouse; murine;
KW Flt3.
XX
OS Mus sp.
XX
PN WO2003039484-A2.
XX
PD 15-MAY-2003.
XX

PF 08-NOV-2002; 2002WO-US036071.
XX
PR 08-NOV-2001; 2001US-00052482.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW, Engelhard EK;
XX
XX WPI; 2003-441462/41.
XX
XX New carcinoma associated nucleic acids and proteins, useful for screening
PT drug candidates, or for diagnosing and treating carcinomas, e.g.
PT lymphoma, breast cancer, prostate cancer or leukemia.
XX
XX Claim 1; SEQ ID NO 133; 793pp; English.
XX
XX This invention relates to novel recombinant nucleic acids for use in
CC diagnosis and treatment of cancer, especially carcinomas, as well as the
CC use of compositions in screening methods. The compositions of the
CC invention may have cytostatic activity whilst the disclosed sequences may
CC be useful for gene therapy. The carcinoma associated nucleic acids and
CC proteins are useful for diagnosing and treating carcinomas, for example
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC drug candidates or bioactive agents capable of binding to, or modulating
CC the activity of, a carcinoma associated protein. The present sequence is
CC the genomic DNA sequence of the mouse Flt3 gene which is a carcinoma
CC associated gene of the invention.
XX
SQ Sequence 47115 BP; 11483 A; 10181 C; 10539 G; 13012 T; 0 U; 1900 Other;
Query Match 73.6%; Score 18.4; DB 10; Length 47115;
Best Local Similarity 95.0%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTC 24
Db 34115 AAGAGCGCAGACTGCTCTTC 34134

RESULT 33
ADZ12547
ID ADZ12547 standard; DNA; 92099 BP.
XX
AC ADZ12547;
XX
XX 16-JUN-2005 (first entry)
DT
DE Murine cancer-associated genomic DNA #8.
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW cytostatic; gene; ds.
XX
OS Mus sp.
XX
XX WO2005031001-A2.
PN
XX 07-APR-2005.
PD
XX 23-SEP-2004; 2004WO-US031617.
PF
XX 23-SEP-2003; 2003US-00669920.
PR
XX (CHIR ) CHIRON CORP.
PA
XX Morris DW, Malandro MS;
PI
XX WPI; 2005-273395/28.
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
PT
XX Disclosure; SEQ ID NO 67; 198pp; English.
XX
```

CC The invention relates to a nucleic acid array for detecting a cancer  
 CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
 CC The invention also relates to a peptide array comprising two or more  
 CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
 CC that binds to a polypeptide, an isolated antibody or its fragment which  
 CC binds to a polypeptide, which is prepared by immunizing a host animal  
 CC with a composition comprising the polypeptide or its antigen binding  
 CC fragment and collecting cells from the host expressing antibodies against  
 CC the antigen or its antigen binding fragment, a composition comprising the  
 CC antibody and a carrier, a method of screening for anticancer activity, a  
 CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
 CC method of treating cancer and a method of inhibiting expression of a CA  
 CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
 CC nucleic acids. The antibody is useful for detecting the presence or  
 CC absence of cancer cells in an individual which involves contacting cells  
 CC from the individual with the antibody and detecting a complex of a CA  
 CC protein from the cancer cells and the antibody, where the detection of  
 CC the complex correlates with the presence of cancer cells in the  
 CC individual. The composition is useful for inhibiting growth of cancer  
 CC cells in an individual or for delivering a therapeutic agent to cancer  
 CC cells in an individual. The invention is also useful for diagnosing  
 CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
 CC a cell. This sequence represents murine cancer-associated genomic DNA of  
 CC the invention.

SQ Sequence 92099 BP; 23678 A; 20842 C; 20864 G; 25593 T; 0 U; 1122 Other;

Query Match 73.6%; Score 18.4; DB 14; Length 92099;  
 Best Local Similarity 95.0%; Pred. No. 2.5e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTC 24  
 DB 78534 AAGAGCGCAGACTGGTCTTC 78553  
 |||||

RESULT 34  
 ABZ55378/c  
 ID ABZ55378 standard; cDNA; 337 BP.  
 XX AC ABZ55378;  
 XX DT 28-MAR-2003 (first entry)  
 XX DE Aspergillus oryzae polynucleotide SEQ ID NO 4491.  
 XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;  
 XX KW expressed sequence tag; gene; ss.  
 XX OS Aspergillus oryzae.

XX WO200279476-A1.  
 XX 10-OCT-2002.  
 XX 22-MAR-2002; 2002WO-IB000890.  
 XX 30-MAR-2001; 2001JP-00098371.  
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX (NARE-) NAT RES INST BREWING.  
 XX (NORQ) NAT FOOD RES INST MIN AGRIC.  
 XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;  
 XX Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;  
 XX WPI; 2003-046817/04.  
 XX Detection of expression of specific Aspergillus genes for monitoring the  
 XX fermentation and growth conditions of the fungus, using DNA probes.

Claim 1; SEQ ID NO 4491; 48pp + Sequence Listing; Japanese.

XX

CC The invention relates to a polynucleotide having any of 6006 specific  
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under  
 CC specific culture conditions including one or more of eutrophic,  
 CC oligotrophic, solid, early germination, alkaline, high temperature, low  
 CC temperature or mold culture or polynucleotides stringently hybridising  
 CC to these sequences. The polynucleotides are useful for monitoring the  
 CC progress of fermentation and the growth conditions of a fungus, trial  
 CC especially of Aspergillus oryzae which is widely used in industrial  
 CC fermentation. Also monitoring for fungal contamination. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 337 BP; 76 A; 68 C; 73 G; 120 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 8; Length 337;  
 Best Local Similarity 87.0%; Pred. No. 1.1e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGCAGACTGGTCTTC 24  
 DB 77 AAAAAGAGCGCTGACAGGTCTTC 55  
 |||||

RESULT 35  
 ADQ97843/c  
 ID ADQ97843 standard; DNA; 92861 BP.  
 XX AC ADQ97843;  
 XX DT 07-OCT-2004 (first entry)  
 XX DE Mouse cancer associated sequence MD11-010, SEQ ID 820.  
 XX KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.

XX Mus musculus.

XX WO2004060304-A2.  
 XX 22-JUL-2004.  
 XX 22-DEC-2003; 2003WO-US041389.  
 XX 27-DEC-2002; 2002US-00330773.  
 XX (SAGR-) SAGRES DISCOVERY INC.  
 XX Morris DW, Malandro MS;  
 XX WPI; 2004-543781/52.

XX New isolated cancer associated nucleic acids comprising at least 10  
 XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
 XX cancers such as leukemia and lymphoma.  
 XX Claim 1; SEQ ID NO 820; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-  
 XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
 XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
 XX data for this patent did not form part of the printed specification, but  
 XX was obtained in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 92861 BP; 25200 A; 20454 C; 20725 G; 24205 T; 0 U; 2277 Other;  
 Query Match 72.8%; Score 18.2; DB 12; Length 92861;  
 Best Local Similarity 87.0%; Pred. No. 3.2e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAGAGCGCAGACTGGTCTTC 25  
 |||||

Db 1111 AAAAGAGCACTGACTGCTCTTCC 1089

RESULT 36  
ADX04494  
ID ADX04494 standard; RNA; 95 BP.  
XX  
AC ADX04494;  
DT 05-MAY-2005 (first entry)  
XX  
DE Rat primary-microRNA (pri-miRNA) mir-333.  
XX  
XX MicroRNA; ds; RNA interference; Gene silencing; Cytostatic; Antidiabetic;  
KW Anorectic; Antilipemic; Antiarteriosclerotic; Hypotensive;  
KW Neuroprotective; Nootropic; Antiangiogenic; Anabolic;  
KW Eating-disorders-gen.; hyperproliferation; cancer; neoplasm;  
KW angiogenesis disorder; cardiovascular disease;  
KW non-insulin dependent diabetes; endocrine disease;  
KW gastrointestinal disease; metabolic disorder; obesity;  
KW nutritional disorder; hyperlipidemia; atherosclerosis; atherogenesis;  
KW hypertension; anorexia nervosa; nutritional disorder;  
KW psychiatric disorder; Alzheimers disease; degeneration;  
KW neurological disease; nervous system injury; neurodegenerative disease;  
KW neurological disorder.  
XX  
OS Rattus sp.  
XX  
XX WO2005013901-A2.  
XX  
XX 17-FEB-2005.  
XX  
XX 30-JUL-2004; 2004WO-US025300.  
XX  
XX 31-JUL-2003; 2003US-0492056P.  
PR 31-OCT-2003; 2003US-0516303P.  
PR 19-DEC-2003; 2003US-0531596P.  
PR 14-APR-2004; 2004US-0562417P.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Esau C, Lollo B, Bennett CF, Freier SM, Griffey RH, Baker BP;  
PI Vickers T, Marcussen EG, Koller E, Swayze EE, Jain R, Bhat B;  
PI Peralta E;  
XX  
XX WPI; 2005-163123/17.  
XX  
XX New oligomeric compound that can hybridize with or sterically interfere  
PT with nucleic acid molecules comprising or encoding small non-coding RNA  
PT targets, useful for treating e.g., cancer and diabetes.  
XX  
XX Example 34; SEQ ID NO 1605; 854pp; English.  
XX  
XX The invention relates to an oligomeric compound comprising a first region  
CC and a second region, where at least one region contains a modification,  
CC and a portion of the oligomeric compound is targeted to a small non-  
CC coding RNA target nucleic acid that is miRNA (micro-RNA), or its  
CC precursor (primary-miRNA, pri-miRNA). Also included are a composition  
CC comprising a first oligomeric compound and a second oligomeric compound  
CC (where at least one of the oligomeric compounds contains a modification,  
CC at least a portion of the first oligomeric compound is capable of  
CC hybridizing with at least a portion of the second oligomeric compound,  
CC and at least a portion of the first oligomeric compound is targeted to a  
CC small non-coding RNA target nucleic acid), a pharmaceutical composition  
CC comprising the composition cited above (and a carrier), a kit or assay  
CC device comprising the composition, modulating the expression of a small  
CC non-coding RNA target nucleic acid in a cell (or tissue or animal),  
CC treating or preventing a disease or disorder associated with a small non-  
CC coding RNA target nucleic acid, treating a condition in an animal,  
CC treating or preventing a disease or disorder associated with CD3e,  
CC methods of screening an oligomeric compound for an effect on miRNA  
CC signaling, methods of screening a miRNA precursor for an effect in miRNA  
CC signaling, methods of modulating translation (or apoptosis, conversion of

CC a precursor miRNA into miRNA, or cellular differentiation), identifying  
CC an RNA transcript bound to a small non-coding RNA, arresting (or  
CC delaying) entry of a cell at the G2/M phase, interfering with chromosome  
CC segregation, a method of triggering apoptosis, detecting a miRNA  
CC precursor, identifying a miRNA target, modulating cellular  
CC differentiation, treating a condition associated with adipocyte  
CC differentiation in an animal, treating/preventing a disease/disorder  
CC associated with aberrant regulation of the cell cycle by miRNAs,  
CC maintaining a pluripotent stem cell and identifying a small non-coding  
CC RNA binding site. The oligomeric compound is targeted to a region  
CC flanking a Drosha cleavage site within a pri-miRNA. It stimulates an  
CC increase in expression of a pri-miRNA. The compounds and compositions are  
CC useful for treating a disease or disorder resulting from chromosomal non-  
CC disjunction, altered methylation, acetylation, or pseudouridylation state  
CC of chromosomes, such as a hyperproliferative condition (e.g. cancer,  
CC neoplasia or angiogenesis), diabetes (Type 2 diabetes), obesity,  
CC hyperlipidemia, atherosclerosis, atherogenesis, hypertension, anorexia,  
CC Alzheimer's disease, a central nervous system injury or neurodegenerative  
CC disorder. The present sequence is a primary miRNA of the invention.  
XX  
SQ Sequence 95 BP; 18 A; 22 C; 32 G; 0 T; 23 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 14; Length 95;  
Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
Matches 15; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AAGAGCGCAGACTGCTCTTCC 25  
Db 61 AAGAGCACAGACUGCUCUCC 81  
RESULT 37  
ABS69109  
ID ABS69109 standard; DNA; 331 BP.  
XX  
AC ABS69109;  
XX  
DT 21-NOV-2002 (first entry)  
XX  
DE Novel murine polynucleotide isolated using gene trap technology #172.  
XX  
KW Mouse; gene trapped sequence; GTS; functional genomic analysis;  
KW phase display system; gene chip; temporal gene expression;  
KW tissue specific gene expression; antisense inhibition; gene targeting;  
KW development disorder; cell differentiation disorder; aging; cancer;  
KW autoimmune disease; lupus; inflammatory disorder; skin disorder;  
KW degenerative disorder; ds.  
XX  
XX Mus musculus.  
XX  
XX US2002102543-A1.  
XX  
XX 01-AUG-2002.  
XX  
XX 30-NOV-2000; 2000US-00728445.  
PF  
XX 01-DEC-1999; 99US-0168358P.  
PR  
XX (FRIE/) FRIEDRICH G.  
PA (ZAMB/) ZAMBROWICZ B.  
PA (SAND/) SANDS A T.  
XX  
XX Friedrich G, Zambrowicz B, Sands AT;  
PI WPI; 2002-690598/74.  
DR  
XX Novel murine polynucleotides that individually identify novel genes into  
PT which a retroviral gene trap vector has integrated, useful in genomic  
PT analysis and in discovery, development of therapeutic and diagnostic  
PT agents.  
XX  
XX Claim 1; Page 75; 296pp; English.  
XX

CC The invention describes an isolated murine polynucleotide (I) comprising  
CC a contiguous stretch of at least 60 nucleotides of one of 265-677  
CC nucleotide 891 OMNIBANK gene trapped sequences (GTSs) (S), given in the  
CC specification. The novel genes and cells are useful in functional genomic  
CC analysis and in the discovery and development of new therapeutic and  
CC diagnostic agents and methods. (I) is useful for identifying the coding  
CC regions of the murine genome, to isolate cDNAs, genomic clones, or full-  
CC length genes/polynucleotides or homologues, heterologues, paralogues, or  
CC orthologues that are capable of hybridising to one or more of the GTSs  
CC under stringent conditions. (I) can be incorporated into a phage display  
CC system that can be used to screen for proteins, or other ligands, that  
CC are capable of binding an amino acid sequence encoded by an  
CC oligonucleotide or polynucleotide sequence in at least one of the TS  
CC sequences. (I) is useful in addressable arrays, such as gene chips, to  
CC identify and characterise temporal and tissue specific gene expression,  
CC to identify the gene of interest from many sources and for genetic  
CC manipulations such as antisense inhibition and gene targeting. Decreasing  
CC the level of expression of (I) and/or down regulating the activity of  
CC peptides or proteins encoded by (I) is useful for treating development  
CC and cell differentiation disorders, aging, cancer, autoimmune disease,  
CC lupus, inflammatory disorders, skin disorders and degenerative disorders.  
CC This sequence represents a murine cDNA isolated using gene trap  
CC technology  
XX  
SQ Sequence 331 BP; 70 A; 79 C; 82 G; 92 T; 0 U; 8 Other;

Query Match 71.2%; Score 17.8; DB 6; Length 331;  
Best Local Similarity 90.5%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
||||| ||||| ||||| ||||| |||||  
Db 184 AAGAGCACAGACTGCTTCTCC 204

RESULT 38  
ADO36057  
ID ADO36057 standard; DNA; 3486 BP.  
XX  
AC ADO36057;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Novel mouse gene sequence #730.  
XX  
KW mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;  
KW ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;  
KW viral disorder; ds; gene.  
XX  
OS Mus sp.  
XX  
PN WO2004046310-A2.  
XX  
PD 03-JUN-2004.  
XX  
PF 24-OCT-2003; 2003WO-US033948.  
XX  
PR 15-NOV-2002; 2002US-0426916P.  
PR 04-DEC-2002; 2002US-0431158P.  
PR 05-DEC-2002; 2002US-0431445P.  
PR 05-DEC-2002; 2002US-0431606P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-AUG-2003; 2003US-0485359P.  
PR 08-AUG-2003; 2003US-0493332P.  
PR 08-AUG-2003; 2003US-0493356P.

XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX  
XX Williams LT, Chu K, Lee E, Hestir K, Hayashizaki Y, Kamiya M;  
XX WPI; 2004-431966/40.  
DR

XX New mouse nucleic acid molecules and polypeptides, useful for treating  
PT cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart  
PT disease or thrombosis.  
XX  
XX Claim 1; SEQ ID NO 730; 263pp; English.  
XX  
XX The invention comprises 744 novel mouse DNA sequences (genes). The DNA  
CC sequences of the invention are useful for treating cancer, psoriasis,  
CC ulcerative colitis, inflammation, ischaemic heart disease, thrombosis,  
CC immune disorders, bacterial disorders and viral disorders. The present  
CC nucleic acid represents a mouse DNA sequence of the invention. NOTE: The  
CC present DNA sequence is not shown in the specification, but has been  
CC retrieved from the WIPO website.  
XX  
SQ Sequence 3486 BP; 836 A; 924 C; 796 G; 929 T; 0 U; 1 Other;

Query Match 71.2%; Score 17.8; DB 12; Length 3486;  
Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
||||| ||||| ||||| ||||| |||||  
Db 2819 AAGAGCACAGACTGCTTCTCC 2839

RESULT 39  
ADO36057/C  
ID ADO36057 standard; DNA; 3486 BP.  
XX  
AC ADO36057;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Novel mouse gene sequence #730.  
XX  
KW mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;  
KW ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;  
KW viral disorder; ds; gene.  
XX  
OS Mus sp.  
XX  
PN WO2004046310-A2.  
XX  
PD 03-JUN-2004.  
XX  
PF 24-OCT-2003; 2003WO-US033948.  
XX  
PR 15-NOV-2002; 2002US-0426916P.  
PR 04-DEC-2002; 2002US-0431158P.  
PR 05-DEC-2002; 2002US-0431445P.  
PR 05-DEC-2002; 2002US-0431606P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-AUG-2003; 2003US-0485359P.  
PR 08-AUG-2003; 2003US-0493332P.  
PR 08-AUG-2003; 2003US-0493356P.

XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX  
XX Williams LT, Chu K, Lee E, Hestir K, Hayashizaki Y, Kamiya M;  
XX WPI; 2004-431966/40.  
XX  
XX New mouse nucleic acid molecules and polypeptides, useful for treating  
PT cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart  
PT disease or thrombosis.  
XX  
XX Claim 1; SEQ ID NO 730; 263pp; English.  
XX  
XX The invention comprises 744 novel mouse DNA sequences (genes). The DNA  
CC sequences of the invention are useful for treating cancer, psoriasis,  
CC ischaemic heart disease, thrombosis, immune disorder; bacterial disorder;  
CC disease or thrombosis.

CC ulcerative colitis, inflammation, ischaemic heart disease, thrombosis,  
CC immune disorders, bacterial disorders and viral disorders. The present  
CC nucleic acid represents a mouse DNA sequence of the invention. NOTE: The  
CC present DNA sequence is not shown in the specification, but has been  
CC retrieved from the WIPO website.

XX  
SQ Sequence 3486 BP; 836 A; 924 C; 796 G; 929 T; 0 U; 1 Other;  
Query Match 71.2%; Score 17.8; DB 12; Length 3486;  
Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
|||||  
DB 164 AAGAGCGCAGCTGCTCTTCC 144

RESULT 40  
ADB81347  
ID ADB81347 standard; DNA; 4333 BP.  
XX  
AC ADB81347;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Murine downstream homology arm inserted into the ROSA26 targeting vector.

XX ds; phic31 integrase; ROSA26 promoter; codon optimised;  
KW site specific recombinase; SSR; gene function; disease model;  
KW gene therapy; transgenic; C31-Int; mouse; murine; homology arm.

XX Mus sp.  
XX WO2003066867-A2.  
PN  
XX 14-AUG-2003.  
XX  
XX 05-FEB-2003; 2003WO-EP001122.  
PF  
XX 06-FEB-2002; 2002US-0354741P.  
PR  
XX (ARTE-) ARTEMIS PHARM GMBH.  
PA Andreas S, Faust N;  
XX  
XX WPI; 2003-663599/62.  
DR  
XX New genetically engineered nucleic acid molecule, useful for preparing an  
PT agent for recombining a DNA molecule containing phic31 integrase  
PT recognition sequences in a eukaryotic cell, a vertebrate or transgenic  
PT organism.

XX Example 3; Page 71-73; 87pp; English.  
PS  
XX This invention relates to novel genetically engineered nucleic acid  
CC molecules encoding phic31 integrase (C31-Int), which has been codon  
CC optimised for expression in eukaryotic host cells. The phic31 integrase  
CC is a site specific recombinase (SSR) that catalyzes recombination between  
CC two phic31 recognition sequences. The introduction of silent mutations  
CC into the coding sequence changes the given codon to one that is most  
CC frequently used in the respective host, which in turn alters expression  
CC levels. Accordingly, using this ability to generate controlled and  
CC permanent modifications in eukaryotic genomes has various research  
CC applications including the study of gene function and the creation of  
CC disease models, as well as gene therapy for medical applications, and the  
CC design of economically important animals and crops. Furthermore, the  
CC phic31 integrase of the invention is useful for preparing an agent for  
CC recombining a DNA molecule containing phic31 integrase recognition  
CC sequences in a eukaryotic cell, a vertebrate or transgenic organism. This  
CC polynucleotide sequence is the murine downstream homology arm located in  
CC the ROSA26 promoter targeting vectors and required for homologous  
CC recombination in ES cells, used in an exemplification of the invention.

SQ Sequence 4333 BP; 1177 A; 814 C; 860 G; 1482 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 10; Length 4333;  
Best Local Similarity 90.5%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
|||||  
DB 1308 AAGAGCACAGACTGCTCTTCC 1328

RESULT 41  
ADQ76467  
ID ADQ76467 standard; DNA; 4333 BP.  
XX  
AC ADQ76467;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Nucleotide sequence of a Rosa26 targeting vector 3' arm.  
XX targeted transgenesis; Rosa26 locus; homologous recombination;  
KW transgenic eukaryotic cell; gene function study; drug development;  
KW disease model; ss.  
XX Mus musculus.

XX EP1439234-A1.  
PN  
XX 21-JUL-2004.  
PD  
XX 08-JAN-2003; 2003EP-00000249.  
PF  
XX 08-JAN-2003; 2003EP-00000249.  
PR  
XX (ARTE-) ARTEMIS PHARM GMBH.

XX Schwenk F, Seibler J, Faust N, Kuehn R;  
PI  
XX WPI; 2004-509160/49.  
DR  
XX Generating transgenic eukaryotic cells having a modified Rosa26 locus  
PT comprises introducing a functional DNA sequence into the Rosa26 locus of  
PT starting eukaryotic cells by homologous recombination with a targeting  
PT vector.

XX Disclosure; SEQ ID NO 5; 26pp; English.  
PS  
XX The specification describes a method for targeted transgenesis, which  
CC generates transgenic eukaryotic cells having a modified Rosa26 locus. The  
CC method comprises introducing a functional DNA sequence into the Rosa26  
CC locus of starting eukaryotic cells by homologous recombination with a  
CC targeting vector comprising the functional DNA sequence flanked by DNA  
CC sequences homologous to the Rosa26 locus. Rosa26 supports the  
CC preservation of inherent activity of heterologous promoters inserted  
CC through homologous recombination into the locus. The method of the  
CC invention is useful for generating transgenic eukaryotic cells. The  
CC eukaryotic cells are useful for gene function studies, drug development  
CC and as disease models. ADQ76466-ADQ76467 represent 5' and 3' arms of a  
CC Rosa26 targeting vector. These two sequences enable homologous  
CC recombination into the Rosa locus.

XX Sequence 4333 BP; 1177 A; 814 C; 860 G; 1482 T; 0 U; 0 Other;  
SQ  
Query Match 71.2%; Score 17.8; DB 12; Length 4333;  
Best Local Similarity 90.5%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
|||||  
DB 1308 AAGAGCACAGACTGCTCTTCC 1328

```
RESULT 42
ADQ82251
ID ADQ82251 standard; DNA; 4333 BP.
XX
AC ADQ82251;
XX
DT 21-OCT-2004 (first entry)
XX
DE Rosa26 locus 3' arm.
XX
KW transgenic eukaryotic cell; Rosa26 locus; gene function study;
KW drug development; ds.
XX
OS Mus sp.
XX
FH Key
FH Location/Qualifiers
FT 1.1616
FT /*tag= a
FT /note= "CAGGS-promoter"
FT polyA_site
FT 3921..4099
FT /*tag= b
FT /note= "CreTR site"
XX
PN WO2004063381-A1.
XX
PD 29-JUL-2004.
XX
XX 08-JAN-2004; 2004WO-EP000065.
XX
XX 08-JAN-2003; 2003EP-00000249.
XX 18-JAN-2003; 2003US-0439367P.
XX
XX (ARTE-) ARTEMIS PHARM GMBH.
XX
XX Schwenk F, Seibler J, Faust N, Kuehn R;
XX WPI; 2004-544090/52.
XX
XX Generating transgenic eukaryotic cells and animals having a modified
PT Rosa26 locus by introducing a functional DNA sequence into the Rosa26
PT locus, useful in gene function studies and drug development.
XX
XX Claim 6; SEQ ID NO 5; 53pp; English.
XX
XX The present invention relates to generating transgenic eukaryotic cells
CC having a modified Rosa26 locus comprising introducing a functional DNA
CC sequence into the Rosa26 locus of starting eukaryotic cells, where the
CC functional DNA sequence is a gene expression cassette comprising a gene
CC of interest operatively linked to a promoter or is a DNA sequence which
CC can be converted into such a gene expression cassette. The methods and
CC compositions of the present invention are useful in targeted transgenesis
CC using the Rosa26 locus. The transgenic cells and organisms are useful for
CC gene function studies, drug development and as disease model animals. The
CC present sequence represents the Rosa26 locus 3' arm.
XX
XX Sequence 4333 BP; 1177 A; 814 C; 860 G; 1482 T; 0 U; 0 Other;
XX
XX Query Match 71.2%; Score 17.8; DB 13; Length 4333;
XX Best Local Similarity 90.5%; Pred. No. 2.8e+02;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 5 AAGAGCGCAGACTGGTCTTCC 25
XX ||||| ||||| |||||
XX Db 1308 AAGAGCACAGACTGCTCTTCC 1328
XX
XX RESULT 44
XX ADT71363
XX ID ADT71363 standard; DNA; 4333 BP.
XX
XX AC ADT71363;
XX
XX DT 16-DEC-2004 (first entry)
XX
XX DE 3' arm Rosa26.
XX
XX Rosa26; ds; mouse; homologous recombination; detectable marker;
XX fluorescence marker; positive selection marker; first detectable marker;
XX RRS; recombinase recognition site; optical selection;
XX second detectable marker.
XX
XX OS Mus musculus.
XX
XX WO2004070040-A1.
XX
XX PD 19-AUG-2004.
XX
XX
XX Murine Rosa26 gene 3' arm.
XX
XX ds; mouse; Rosa26; targeted homologous recombination; gene; transgenic.
XX
XX Mus musculus.
XX
XX EP1445320-A1.
XX
XX 11-AUG-2004.
XX
XX 05-FEB-2003; 2003EP-00002424.
XX
XX 05-FEB-2003; 2003EP-00002424.
XX (ARTE-) ARTEMIS PHARM GMBH.
XX
XX Kauselmann G, Zevnik B, Seibler J, Kern H;
XX WPI; 2004-563471/55.
XX
XX Novel vector comprising homologous targeting cassette harboring
PT functional DNA sequence, and expression cassette harboring DNA sequence
PT coding for first detectable marker, useful for targeted homologous
PT recombination of eukaryotic cells.
XX
XX Example 1; SEQ ID NO 5; 46pp; English.
XX
XX The present invention relates to a vector for targeted homologous
CC recombination of eukaryotic cells, comprising a homologous targeting
CC cassette harbouring a functional DNA sequence, and an expression cassette
CC harbouring a functional DNA sequence coding for a first detectable
CC marker, with the expression cassette being connected with the homologous
CC targeting cassette so as to allow distinction between targeted and non-
CC targeted cells. The vector is useful for homologous recombination of
CC eukaryotic cells, which is useful for preparing transgenic tissues,
CC organs and/or multi-cell organisms. The transgenic multi-cell organism is
CC a non-human mammal, most preferably mouse. The present sequence is a
CC fragment of the murine Rosa26 gene used in the exemplification of the
CC invention.
XX
XX Sequence 4333 BP; 1177 A; 814 C; 860 G; 1482 T; 0 U; 0 Other;
XX
XX Query Match 71.2%; Score 17.8; DB 13; Length 4333;
XX Best Local Similarity 90.5%; Pred. No. 2.8e+02;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 5 AAGAGCGCAGACTGGTCTTCC 25
XX ||||| ||||| |||||
XX Db 1308 AAGAGCACAGACTGCTCTTCC 1328
XX
XX RESULT 44
XX ADT71363
XX ID ADT71363 standard; DNA; 4333 BP.
XX
XX AC ADT71363;
XX
XX DT 16-DEC-2004 (first entry)
XX
XX DE 3' arm Rosa26.
XX
XX Rosa26; ds; mouse; homologous recombination; detectable marker;
XX fluorescence marker; positive selection marker; first detectable marker;
XX RRS; recombinase recognition site; optical selection;
XX second detectable marker.
XX
XX OS Mus musculus.
XX
XX WO2004070040-A1.
XX
XX PD 19-AUG-2004.
XX
```



PF 03-FEB-2004; 2004WO-EP000968.  
XX  
XX  
PR 05-FEB-2003; 2003EP-00002424.  
PR 12-FEB-2003; 2003US-0446921P.  
XX  
PA (ARTE-) ARTEMIS PHARM GMBH.  
XX  
XX  
PI Kauselmann G, Zevnik B, Seibler J, Kern H;  
XX WPI; 2004-604456/58.  
XX  
PT Isolating primary mammalian after homologous recombination comprises  
PT transfecting primary cells with a homologous targeting vector comprising  
PT cassette with detectable marker.  
XX  
XX  
PS Example 1; SEQ ID NO 5; 67pp; English.  
XX  
CC The invention provides the method for automatic identification and  
CC isolation of cells harbouring a predetermined genetic modification  
CC (homologous recombination) by using a detectable marker (fluorescence  
CC marker). The homologous targeting sequence comprises a functional DNA  
CC segment and a positive selection marker, or a functional DNA segment and  
CC a first detectable marker and two DNA segments homologous to the  
CC integration site within the genome. The second expression cassette  
CC comprises the DNA sequence coding for a second detectable marker which is  
CC different from first detectable marker. The gene targeting vector  
CC contains a positive selection marker (neomycin) inside the targeting  
CC cassette and a fluorescent marker outside the region of homology. The  
CC positive selection marker is flanked by one or more RRSs (recombinase  
CC recognition sites) and encoding a protein which is resistance against  
CC cell poison. The first and second detectable markers are DNA sequences  
CC encoding a protein which allows direct or indirect optical detection. The  
CC second detectable marker differs from the first detectable marker to  
CC allow separate detection. The detectable marker is a membrane bound or  
CC protein with membrane anchoring signal sequence. The vectors for targeted  
CC homologous recombination contain one or more expression cassette(s)  
CC coding for a detectable marker placed outside the targeting cassette,  
CC which allows rapid and reliable distinction between targeted and non-  
CC targeted ES cells in an automated fashion. The detection strategy of the  
CC invention allows selective identification of the gene targeted embryonic  
CC stem cell clones without exposing the ES cells to potentially toxic  
CC chemicals, proteins or mutagens. It provides automated optical selection  
CC of clones enriched for a homologous recombination event by applying  
CC gentle sorting machinery. This method identifies a higher percentage of  
CC clones showing homologous recombination compared to the standard  
CC procedure. This method also reduces the manual labour and costs (required  
CC for gene targeting experiment). The first detectable marker in a  
CC preferred vector of the invention is ZeGreen and the constitutive  
CC promoter is the CAGGS promoter. The Rosa locus of the mouse was chosen  
CC for homologous recombination. The methods are useful for isolating  
CC primary mammalian cells after homologous recombination and for preparing  
CC transgenic tissues, organs and multi-cell organisms. The presented  
CC nucleotide sequence is the mouse 3' arm of Rosa26.  
XX  
SQ Sequence 4333 BP; 1177 A; 814 C; 860 G; 1482 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 13; Length 4333;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AAGAGCGCAGACTGGCTCTCC 25  
Db 1308 AAGAGCACAGACTGCTCTTCC 1328  
RESULT 45  
ID ADN11341  
XX ADN11341 standard; DNA; 5409 BP.  
AC ADN11341;  
XX  
XX 15-JUL-2004 (first entry)  
XX

DE Murine Rosa26 locus homology region sequence SEQ ID NO:2.  
XX  
XX expression vector; short hairpin RNA; shRNA;  
KW constitutive gene knock down; inducible gene knock down;  
KW polymerase II dependent locus; gene therapy; murine; Rosa26 locus; gene;  
KW ds.  
XX  
XX Mus sp.  
OS Synthetic.  
OS  
PN WO2004035782-A2.  
XX  
PD 29-APR-2004.  
XX  
XX 10-OCT-2003; 2003WO-EP011233.  
XX  
XX 17-OCT-2002; 2003EP-00023283.  
PR 02-MAY-2003; 2003US-0467814P.  
PR 10-JUL-2003; 2003US-0485969P.  
XX  
XX (ARTE-) ARTEMIS PHARM GMBH.  
PA  
PI Seibler J, Schwenk F, Kuehn R, Kueter-Luks B;  
XX WPI; 2004-348459/32.  
DR  
XX New expression vector comprising a short hairpin RNA (shRNA) construct  
PT under the control of a ubiquitous promoter, useful for constitutive  
PT and/or inducible gene knock down in a vertebrate, or in a tissue.  
XX  
PS Example; SEQ ID NO 2; 70pp; English.  
XX  
CC The present invention describes an expression vector comprising a short  
CC hairpin RNA (shRNA) construct under the control of a ubiquitous promoter.  
CC Also described: (1) a method for constitutive and/or inducible gene knock  
CC down in a vertebrate, or in a tissue culture or cells of a cell culture  
CC derived from a vertebrate by stably integrating an expression vector into  
CC the genome of the vertebrate, of the tissue culture or of the cells of  
CC the cell culture; and (2) a vertebrate or tissue or cell culture derived  
CC from a vertebrate having stably integrated, preferably at a polymerase II  
CC dependent locus or the vertebrate, tissue culture or cells of the tissue  
CC culture, an expression vector as defined above. The expression vector can  
CC be used in gene therapy. The expression vector is useful for preparing an  
CC agent, shRNA, for constitutive and/or inducible gene knock down in a  
CC vertebrate, or in a tissue culture or cells of a cell culture derived  
CC from a vertebrate. The present sequence represents a nucleotide sequence  
CC which is used in the exemplification of the present invention.  
XX  
SQ Sequence 5409 BP; 1332 A; 1123 C; 1282 G; 1672 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 12; Length 5409;  
Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AAGAGCGCAGACTGGCTCTCC 25  
Db 2384 AAGAGCACAGACTGCTCTTCC 2404  
RESULT 46  
ID ADB81348  
XX ADB81348 standard; DNA; 10491 BP.  
AC ADB81348;  
XX  
XX 04-DEC-2003 (first entry)  
XX  
XX PROSA12 vector comprising an adenoviral splice acceptor.  
XX  
XX ROSA26 promoter; ds; phiC31 integrase; proSA12;  
KW site specific recombinase; SSR; gene function; disease model;  
KW gene therapy; transgenic; C31-Int; mouse; murine.  
XX

OS unidentified adenovirus.  
 XX Mus sp.  
 XX  
 PN WO2003066867-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 05-FEB-2003; 2003WO-EP001122.  
 XX  
 PR 06-FEB-2002; 2002US-0354741P.  
 XX  
 XX (ARTE-) ARTEMIS PHARM GMBH.  
 XX  
 XX Andreas S, Faust N;  
 XX  
 PI WPI; 2003-663599/62.  
 XX  
 DR  
 XX  
 XX New genetically engineered nucleic acid molecule, useful for preparing an  
 PT agent for recombining a DNA molecule containing phic31 integrase  
 PT recognition sequences in a eukaryotic cell, a vertebrate or transgenic  
 PT organism.  
 XX  
 XX Example 3; Page 73-79; 87pp; English.  
 PS  
 XX This invention relates to novel genetically engineered nucleic acid  
 XX molecules encoding phic31 integrase (C31-Int), which has been codon  
 CC optimised for expression in eukaryotic host cells. The phic31 integrase  
 CC is a site specific recombinase (SSR) that catalyzes recombination between  
 CC two phic31 recognition sequences. The introduction of silent mutations  
 CC into the coding sequence changes the given codon to one that is most  
 CC frequently used in the respective host, which in turn alters expression  
 CC levels. Accordingly, using this ability to generate controlled and  
 CC permanent modifications in eukaryotic genomes has various research  
 CC applications including the study of gene function and the creation of  
 CC disease models, as well as gene therapy for medical applications, and the  
 CC design of economically important animals and crops. Furthermore, the  
 CC phic31 integrase of the invention is useful for preparing an agent for  
 CC recombining a DNA molecule containing phic31 integrase recognition  
 CC sequences in a eukaryotic cell, a vertebrate or transgenic organism. This  
 CC polynucleotide sequence is the pRosa12 vector comprising an adenoviral  
 CC splice acceptor site inserted between the two ROSA26 promoter homology  
 CC arms, used in an exemplification of the invention.  
 XX  
 XX Sequence 10491 BP; 2499 A; 2433 C; 2608 G; 2951 T; 0 U; 0 Other;  
 SQ  
 Query Match 71.2%; Score 17.8; DB 10; Length 10491;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 AAGAGCGCAGACTGGTCTTCC 25  
 DB 4954 AAGAGCACAGACTGCTCTTCC 4974  
 RESULT 47  
 AAS01080  
 ID AAS01080 standard; DNA; 11176 BP.  
 XX  
 AC AAS01080;  
 XX  
 XX 29-MAY-2001 (first entry)  
 DT  
 XX Mouse FosB DNA.  
 DE  
 XX Mouse; phosphoglycerate kinase 1 promoter; targeting construct;  
 KW transgenic animal; reporter expression cassette; luciferase; vitronectin;  
 KW FosB; galactin 3; VEGF; VEGFR; Tie2; selectable marker;  
 KW neomycin phosphotransferase II; chloramphenicol acetyltransferase;  
 KW thymidine kinase; PCR primer; ds.  
 XX  
 OS Mus sp.  
 XX  
 XX WO2000118225-A1.  
 PN

XX 15-MAR-2001.  
 XX  
 XX 16-DEC-1999; 99WO-US030078.  
 XX  
 PR 03-SEP-1999; 99US-0152522P.  
 XX  
 XX (XENO-) XENGEN CORP.  
 PA  
 XX Zhang N;  
 PI  
 XX WPI; 2001-203085/20.  
 DR  
 XX Targeting constructs and methods of using them for creating transgenic  
 XX animals.  
 PT  
 XX Disclosure; Fig 4B; 96pp; English.  
 PS  
 XX The present sequence for mouse FosB DNA is used to generate a targeting  
 CC cassette and vector in the invention of novel targeting constructs and  
 CC methods of using them for creating transgenic animals in which at least 1  
 CC single-copy, non-essential gene is replaced with a reporter expression  
 CC cassette (e.g. a luciferase gene linked to a promoter heterologous to the  
 CC single-copy, non-essential gene). The single-copy, non-essential gene may  
 CC be selected from vitronectin (Vn), fosB and galactin 3. The promoter  
 CC element may be selected from VEGF, VEGFR and Tie2. The constructs may  
 CC contain a selectable marker e.g. neomycin phosphotransferase II,  
 CC chloramphenicol acetyltransferase and thymidine kinase. The constructs  
 CC are useful for creating transgenic animals in which at least 1 single-  
 CC copy, non-essential gene is replaced with a reporter expression cassette.  
 CC Methods of using these animals are also described  
 XX  
 XX Sequence 11176 BP; 2570 A; 2863 C; 2801 G; 2941 T; 0 U; 1 Other;  
 SQ  
 Query Match 71.2%; Score 17.8; DB 4; Length 11176;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 AAGAGCGCAGACTGGTCTTCC 25  
 DB 10526 AAGAGCACAGACTGCTCTTCC 10546  
 RESULT 48  
 AAD22528  
 ID AAD22528 standard; DNA; 11176 BP.  
 XX  
 XX AAD22528;  
 AC  
 XX 12-FEB-2002 (first entry)  
 DT  
 XX Mouse FosB DNA.  
 DE  
 XX Transcription control element; angiogenesis; tumorigenesis; VEGFR-2;  
 KW transgenic animal; vascular endothelial growth factor; VEGF; Tie2 gene;  
 KW VEGF receptor 2; PCR primer; FosB gene; mouse; ss.  
 XX  
 XX Mus sp.  
 OS  
 XX US2001037016-A1.  
 PN  
 XX 01-NOV-2001.  
 PD  
 XX 15-DEC-2000; 2000US-00738968.  
 PF  
 XX 16-DEC-1999; 99US-00465978.  
 PR  
 XX (NING/) NING Z.  
 PA (CONT/) CONTAG P R.  
 PA (PURC/) PURCHIO A F.  
 XX  
 XX Ning Z, Contag PR, Purchio AF;  
 PI  
 XX

DR WPI; 2002-017156/02.  
XX Polynucleotide transcription control elements derived from vascular  
PT endothelial growth factor, and receptors useful for identifying  
PT modulators useful for modulating tumorigenesis.  
XX Example 1; Page 45-50; 85pp; English.  
PS The invention relates to transcription control elements including  
XX promoters derived from mouse vascular endothelial growth factor (VEGF),  
CC VEGF receptor 2 (VEGFR-2) and Tie2 gene loci. The invention also relates  
CC to expression cassettes and vector constructs incorporating promoters  
CC useful for generating transformed cells and transgenic animals and  
CC methods for screening therapeutic compounds that modulate angiogenesis  
CC and tumorigenesis. The present sequence is mouse FosB DNA used in the  
CC construction of vectors containing VEGF, VEGFR-2 and Tie2 expression  
CC cassettes  
XX  
SQ Sequence 11176 BP; 2570 A; 2863 C; 2801 G; 2941 T; 0 U; 1 Other;  
Query Match 71.2%; Score 17.8; DB 6; Length 11176;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AAGAGCGCAGACTGCTCTTCC 25  
DB 10526 AAGAGCAGACAGCTGCTCTTCC 10546  
RESULT 49  
ABT08201  
ID ABT08201 standard; DNA; 11784 BP.  
AC ABT08201;  
XX 28-NOV-2002 (first entry)  
XX Modified ROSA26 locus.  
DE  
XX Fusion protein; recombinase domain; signal peptide domain; gene function;  
XX nuclear import; recombinase recognition sequence; transgenic organism;  
KW C31-Int recombination system; site-specific integration; gene therapy;  
KW ds.  
XX Unidentified.  
OS Synthetic.  
OS WO200238613-A2.  
PN  
XX 16-MAY-2002.  
PD  
XX 09-NOV-2001; 2001WO-EP012975.  
PF  
XX 10-NOV-2000; 2000EP-00124629.  
PR  
XX 17-APR-2001; 2001EP-00109543.  
PR  
XX 13-AUG-2001; 2001US-0311876P.  
XX  
XX (ARTE-) ARTEMIS PHARM GMBH.  
PA  
XX Kuehn R, Felder S, Schwenk F, Kueter-Luks B, Faust N;  
PI  
XX WPI; 2002-519298/55.  
DR  
XX Novel fusion protein useful for recombining DNA molecules in eukaryotic  
PT cells has recombinase protein which is linked to signal peptide domain  
PT which directs nuclear import of fusion protein in eukaryotic cells.  
PT  
XX Example 4; Page 135-138; 150pp; English.  
PS The invention comprises the amino acid and coding sequences of fusion  
XX proteins that contain a recombinase domain and a signal peptide domain  
CC which directs nuclear import of the fusion protein in eukaryotic cells.  
CC The fusion proteins of the invention are useful for recombining the DNA

CC molecules of cells or organisms containing recombinase recognition  
CC sequences for the recombinase domain of the fusion proteins. The fusion  
CC proteins of the invention are useful for studying gene function at  
CC various developmental stages and for the creation of transgenic  
CC organisms. The C31-Int recombination system of the invention can be used  
CC for the site-specific integration of foreign DNA into the genome of  
CC mammalian cells (e.g. for gene therapy). The present DNA sequence  
CC represents an oligonucleotide that was used in the invention  
XX  
SQ Sequence 11784 BP; 2796 A; 2787 C; 3027 G; 3174 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 6; Length 11784;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 AAGAGCGCAGACTGCTCTTCC 25  
DB 8760 AAGAGCAGACAGCTGCTCTTCC 8780  
RESULT 50  
ADB81350  
ID ADB81350 standard; DNA; 11784 BP.  
XX  
XX ADB81350;  
XX  
XX 04-DEC-2003 (first entry)  
XX Bacteriophage C31 substrate reporter DNA.  
XX ROSA26 promoter; ds; phiC31 integrase; PROSA12;  
KW site specific recombinase; SSR; gene function; disease model;  
KW gene therapy; transgenic; C31-Int.  
XX Bacteriophage phi-C31.  
XX WO2003066867-A2.  
PN  
XX 14-AUG-2003.  
PD  
XX 05-FEB-2003; 2003WO-EP001122.  
PF  
XX 06-FEB-2002; 2002US-0354741P.  
PR  
XX (ARTE-) ARTEMIS PHARM GMBH.  
PA  
XX Andreas S, Faust N;  
PI  
XX WPI; 2003-663599/62.  
DR  
XX New genetically engineered nucleic acid molecule, useful for preparing an  
PT agent for recombining a DNA molecule containing phiC31 integrase  
PT recognition sequences in a eukaryotic cell, a vertebrate or transgenic  
PT organism.  
XX  
XX Example 3; Page 79-86; 87pp; English.  
PS  
XX This invention relates to novel genetically engineered nucleic acid  
CC molecules encoding phiC31 integrase (C31-Int), which has been codon  
CC optimised for expression in eukaryotic host cells. The phiC31 integrase  
CC is a site specific recombinase (SSR) that catalyzes recombination between  
CC two phiC31 recognition sequences. The introduction of silent mutations  
CC into the coding sequence changes the given codon to one that is most  
CC frequently used in the respective host, which in turn alters expression  
CC levels. Accordingly, using this ability to generate controlled and  
CC permanent modifications in eukaryotic genomes has various research  
CC applications including the study of gene function and the creation of  
CC disease models, as well as gene therapy for medical applications, and the  
CC design of economically important animals and crops. Furthermore, the  
CC phiC31 integrase of the invention is useful for preparing an agent for  
CC recombining a DNA molecule containing phiC31 integrase recognition  
CC sequences in a eukaryotic cell, a vertebrate or transgenic organism. This  
CC polynucleotide sequence is the bacteriophage phi-31 substrate reporter

CC DNA located in the ROSA26 locus of heterozygous mice, used in an  
CC exemplification of the invention.  
XX  
SQ Sequence 11784 BP; 2796 A; 2787 C; 3027 G; 3174 T; 0 U; 0 Other;  
Query Match 71.2%; Score 17.8; DB 10; Length 11784;  
Best Local Similarity 90.5%; Pred. NO. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 5 AGAGCGCAGACTGCTCTTCC 25  
Db 8760 AGAGCACAGACTGCTCTTCC 8780

Search completed: February 3, 2006, 21:57:20  
Job time : 210.111 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:38:25 ; Search time 2951.67 Seconds  
(without alignments)  
396.277 Million cell updates/sec

Title: US-10-719-900-35

Perfect score: 25

Sequence: 1 aaaaagagcgagactggtcttcc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_est9.\*

10: gb\_est10.\*

11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	655	1	BB066943
2	23.4	93.6	2333	4	AK033191 Mus muscu
3	21.8	87.2	346	5	BY306264 BY306264
4	21.8	87.2	760	10	AG405809 AG405809 Mus muscu
5	21.8	87.2	938	5	B0925815 B0925815 AGNCOURT
6	20.2	80.8	212	2	BB371564 BB371564
7	20.2	80.8	249	6	CA504477 UI-R-P81-
8	20.2	80.8	330	9	B2192915 CH230-248
9	20.2	80.8	347	1	AU018134 AU018134
10	20.2	80.8	352	9	A2947222 2M0209B21
11	20.2	80.8	358	2	BG071090 H3094F01-
12	20.2	80.8	363	9	BH315230 CH230-159
13	20.2	80.8	371	2	BG083952 H3094F01-
14	20.2	80.8	393	9	BH344887 CH230-183
15	20.2	80.8	400	7	CN669196 A0876H02-
16	20.2	80.8	417	1	AW535711 UI-R-B50-
17	20.2	80.8	420	9	B2207783 CH230-475
18	20.2	80.8	430	5	BY567189 BY567189
19	20.2	80.8	443	3	BP417531 BP417531
20	20.2	80.8	464	9	A2960237 2M0228P09
21	20.2	80.8	480	9	BH260028 CH230-113
22	20.2	80.8	481	5	BY566957 BY566957

482	3	BP775662	BP775662
506	3	BI301386	BI301386
509	9	AZ875623	AZ875623 2M0190D13
528	2	BB756913	BB756913
536	9	BH349233	BH349233 CH230-81B
574	2	BG088328	BG088328 H3151G04-
577	9	AZ288000	AZ288000 RPCI-23-1
586	9	BH261566	BH261566 CH230-90J
607	9	AZ357207	AZ357207 IM0098D08
625	7	CN542163	CN542163 UI-R-EAO-
632	9	BZ116863	BZ116863 CH230-510
702	10	AG282644	AG282644 Mus muscu
708	9	BZ196833	BZ196833 CH230-269
714	10	AG562735	AG562735 Mus muscu
726	2	BG298680	BG298680 602396890
736	9	AZ430769	AZ430769 IM0215G22
779	9	BZ173903	BZ173903 CH230-493
789	9	BH292979	BH292979 CH230-92B
809	9	BZ277088	BZ277088 CH230-308
816	10	AG295893	AG295893 Mus muscu
831	5	BQ959424	BQ959424 AGNCOURT
1052	9	BZ275313	BZ275313 CH230-450
401	1	AA138597	AA138597 mg92C11.f
868	7	CO034709	CO034709 ESF813093
1359	1	AA590658	AA590658 vm24G12.f
277	1	AA823597	AA823597 vw38A05.f
292	11	CR109101	CR109101 Reverse s
343	2	BE197804	BE197804 us77e02.x
347	2	BG094937	BG094937 mac25H02.
356	5	BY210447	BY210447
365	5	BY209501	BY209501 BY209501
375	1	AI413095	AI413095 md92C07.x
382	8	W62939	W62939 md92C07.r1
389	5	W48161	W48161 mc86G06.r1
427	5	BY545965	BY545965 BY545965
436	1	AA822961	AA822961 vw39A05.f
443	1	AI430486	AI430486 md32C07.y
467	5	BY256574	BY256574
475	1	AA690208	AA690208 vr78G11.s
477	1	AI427277	AI427277 mc86e06.y
487	9	AZ017695	AZ017695 RPCI-23-2
494	9	AZ403580	AZ403580 1M0171007
516	1	AA164091	AA164091 mr23d07.f
543	4	AK020460	AK020460 Mus muscu
558	5	BY217120	BY217120 BY217120
564	2	BG609520	BG609520 323403 MA
597	11	CR043665	CR043665 Reverse s
604	7	CN664053	CN664053 A0808B01-
605	9	AZ403386	AZ403386 1M0171M07
609	5	BQ604998	BQ604998 MI-P-CPI-
626	2	BB664023	BB664023 BB664023
635	11	CR258177	CR258177 Forward s
642	5	BY718759	BY718759 BY718759
648	6	CF171469	CF171469 B0842H03-
661	9	AZ839689	AZ839689 2M0135K23
686	5	BX669334	BX669334 BX669334
686	7	CJ017268	CJ017268 CJ017268
718	7	CJ038423	CJ038423 CJ038423
815	3	BP172852	BP172852 BP172852
848	11	CR192365	CR192365 Reverse s
862	11	CR213048	CR213048 Reverse s
1056	8	W83215	W83215 mf23G02.r1
1102	4	AK078159	AK078159 Mus muscu
1761	4	AK087053	AK087053 Mus muscu
193	3	BQ166077	BQ166077 WHE0808.D
390	9	AZ108024	AZ108024 RPCI-23-2
406	9	BZ197313	BZ197313 CH230-416
477	9	AZ613766	AZ613766 IM0442113
548	7	CF898666	CF898666 A0242F05-
579	9	AQ487165	AQ487165 RPCI-11-2
599	9	AQ486429	AQ486429 RPCI-11-2
601	11	CR180552	CR180552 Reverse s
607	9	BH294574	BH294574 CH230-104







Numata, K., Okido, T., Pavan, W. J., Perteau, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

**TITLE** Analysis of the mouse transcriptome based on functional annotation

**JOURNAL** Nature 420, 563-573 (2002)

**PUBLISHED** 12466851

**COMMENT** Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues

Tissues were provided by Michelle Brinkmeier and Sally Camper (

Dept. Human Genetics University of Michigan Medical School 4301

MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA )

whose assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

**FEATURES** Location/Qualifiers

source

1. .346

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CD-1"

/db\_xref="taxon:10090"

/clone="K920009A05"

/tissue\_type="Rathke's pouches"

/dev\_stage="12.5 days embryo"

/clone\_lib="RIKEN full-length enriched, 12.5 days embryo

Rathke's pouches"

**ORIGIN**

Query Match 87.2%; Score 21.8; DB 5; Length 346;

Best Local Similarity 92.0%; Pred. No. 28;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGCTGCTCTTCC 25

|||||

**Db****RESULT 4****LOCUS****DEFINITION****ACCESSION****VERSION****KEYWORDS****SOURCE****ORGANISM****REFERENCE****AUTHORS****TITLE****JOURNAL****PUBLISHED****REFERENCE****AUTHORS****TITLE****JOURNAL****COMMENT****FEATURES****source****ORIGIN**

Query Match 87.2%; Score 21.8; DB 10; Length 760;

Best Local Similarity 92.0%; Pred. No. 32;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGCTGCTCTTCC 25

|||||

Db 576 AAAAAAGCGCAGCTGCTCTTCC 600

**RESULT 5****LOCUS****DEFINITION****ACCESSION**

Query Match 87.2%; Score 21.8; DB 5; Length 346;

Best Local Similarity 92.0%; Pred. No. 28;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

BQ925815

AGENCOURT 8801758 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6315307

5', mRNA sequence.

BQ925815

45 AAAAAAGAGCACTGACTGCTCTTCC 21

**LOCUS****DEFINITION****ACCESSION****VERSION****KEYWORDS****SOURCE****ORGANISM****REFERENCE****AUTHORS****TITLE****JOURNAL****PUBLISHED****REFERENCE****AUTHORS****TITLE****JOURNAL****COMMENT****FEATURES****source****ORIGIN**

Query Match 87.2%; Score 21.8; DB 10; Length 760;

Best Local Similarity 92.0%; Pred. No. 32;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGCTGCTCTTCC 25

|||||

Db 576 AAAAAAGCGCAGCTGCTCTTCC 600

VERSION	BQ925815.1	GI:22340846	TITLE	RIKEN Mouse ESTs (Konno,H., et al.)
KEYWORDS	EST.		JOURNAL	Unpublished (2000)
SOURCE	Mus musculus (house mouse)		COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.jp, URL:http://genome.gsc.riken.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
ORGANISM	Mus musculus			
REFERENCE	1 (bases 1 to 938)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcapbs@mail.nih.gov Tissue Procurement: Susan L. Sullivan, PhD. cDNA Library Preparation: ResGen, Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM13745 row: g column: 20 High quality sequence stop: 625.			
FEATURES	Location/Qualifiers			
source	1..938			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:6315307"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_lib="NIH_MGC_129"			
	/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."			
ORIGIN				
Query Match	87.2%;	Score 21.8;	DB 5;	Length 938;
Best Local Similarity	92.0%;	Pred. No. 33;		
Matches	23;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Oy	1	AAAAAAGAGCGCAGACTGGTCTTCC 25		
Dd	405	AAAAAAGAGCAGTGGTCTTCC 381		
RESULT 6				
BB371564				
LOCUS	BB371564	212 bp	mRNA	linear EST 13-JUL-2000
DEFINITION	BB371564	RIKEN full-length enriched, 16 days embryo head Mus		
ACCESSION	BB371564	musculus cDNA clone C130060G02 3', mRNA sequence.		
VERSION	BB371564.1	GI:9084058		
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 212)			
AUTHORS	Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kuakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigenoto,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Tsumoda,Y., Tagawa,A., Takahashi,F., Tomimaga,N., Toyota,T., Tsumoda,Y., Watanabe,S., Watanabe,T., Yamamura,T., Yamana,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.			
ORIGIN				
Query Match	80.8%;	Score 20.2;	DB 2;	Length 212;
Best Local Similarity	88.0%;	Pred. No. 1.5e+02;		
Matches	22;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
Oy	1	AAAAAAGAGCGCAGACTGGTCTTCC 25		
Dd	61	AAAAAAGAGCAGTGGTCTTCC 85		
RESULT 7				
BB371564				
LOCUS	BB371564	249 bp	mRNA	linear EST 12-AUG-2004
DEFINITION	BB371564	UI-R-PS1-cqe-b-19-0-UI-s1 NCI CGAP_FSI Rattus norvegicus cDNA clone		
ACCESSION	BB371564	IMAGE:7361685 3', mRNA sequence.		
VERSION	BB371564.1	GI:24995431		
KEYWORDS	EST.			

SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 249)  
 AUTHORS Ronaldo M.F., Lennon G. and Soares M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 PUBMED 889548  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics / NIH  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabs-i@mail.nih.gov  
 Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA library Preparation: M.B.  
 Soares Lab Clone distribution: Clones will be available through  
 IMAGE (http://image.llnl.gov)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source  
 1..249  
 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="IMAGE:7361685"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI\_CGAP\_FSI"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not 1; Site 2: Eco RI; UI-R-FS1 is a  
 normalized cDNA library constructed from Swam Rat  
 Chondrosarcoma. The library was constructed and  
 normalized according to Ronaldo, Lennon and Soares, Genome  
 Research, 6:791-806, 1996. The Rat SRC-JWS Cell Line tumor  
 was provided by Dr Jeff Stevens of the University of Iowa.  
 TAG TISSUE=Rat SRC-JWS Cell Line  
 TAG LIB=UI-R-FS1  
 TAG\_SEQ=AGCGCCGAT"

ORIGIN  
 Query Match 80.8%; Score 20.2; DB 6; Length 249;  
 Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTTCC 25  
 |||||  
 Db 173 AAAAAAGCGCAGCTGCTCTTCC 149  
 |||||

RESULT 8  
 BZ192915/c  
 LOCUS Rattus norvegicus 330 bp DNA linear GSS 11-OCT-2002  
 DEFINITION CH230-248022.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 CH230-248022, genomic survey sequence.  
 ACCESSION BZ192915  
 VERSION BZ192915.1 GI:23850972  
 KEYWORDS GSS.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 330)  
 AUTHORS Zhao S., Shetty J., Shatsman S., Teegave G., Geer K.,  
 Shvartsbeyn A., Gebregorgis E., Overton L., Russell D., Chen D.,  
 Riggs F., de Jong P. and Fraser C.M.  
 TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: CH230-248022.TV

Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering information.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 248 row: 0 column: 22  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..330  
 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SENHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-248022"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: pTARBAC1-3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
 Pieter de Jong"

ORIGIN  
 Query Match 80.8%; Score 20.2; DB 9; Length 330;  
 Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTTCC 25  
 |||||  
 Db 258 AAAAAAGCGCTGCTGCTCTTCC 234  
 |||||

RESULT 9  
 AU018134/c  
 LOCUS AU018134 Mouse two-cell stage embryo cDNA Mus musculus EST 19-OCT-1998  
 DEFINITION J0749B12 3', mRNA sequence.  
 ACCESSION AU018134  
 VERSION AU018134.1 GI:3373624  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 347)  
 AUTHORS Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., DePalma, G.E.,  
 Liang, Y., Kargul, G.J., Sharara, R. and Doi, H.  
 TITLE Systematic analyses of genes expressed in 2-cell stage mouse  
 embryos (The ERATO/Doi Project at Wayne State University)  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Hirofumi Doi  
 Doi Biosymmetry Project, ERATO  
 Japan Science and Technology Corporation (JST)  
 WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan  
 Email: hdo@bioa.jst.go.jp.  
 Location/Qualifiers  
 source  
 1..347  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="J0749B12"  
 /dev\_stage="two-cell stage embryo"

/clone\_lib="Mouse two-cell stage embryo cDNA"

## ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 347;  
 Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
 |||||  
 Db 168 AAAAAAGAGCAGCTGCTCTTCC 144

## RESULT 10

AZ947222/c  
 LOCUS AZ947222 352 bp DNA linear GSS 27-APR-2001  
 DEFINITION 2M0209B21F Mouse 10kb plasmid UGCG2M library Mus musculus genomic  
 clone UGCG2M0209B21 F, genomic survey sequence.

## ACCESSION

VERSION AZ947222

## KEYWORDS

GSS.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 352)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,

Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

Niederhausern,A. and Wright,D., Weis,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0209 row: B column: 21

Seq primer: CGTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 352.

Location/Qualifiers

1. .352

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UGCG2M0209B21"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UGCG2M library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 [gi|4732114|gb|AF129072.1], a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 352;  
 Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
 |||||  
 Db 81 AAAAAAGAGCAGCTGCTCTTCC 57

## RESULT 11

BG071090/c  
 LOCUS BG071090 358 bp mRNA linear EST 26-JAN-2001  
 DEFINITION H3094F01-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 H3094F01 3', mRNA sequence.

## ACCESSION

VERSION BG071090

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 358)

Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A.,

Tanaka,T.S., Carter,M.G. and Ko,M.S.H

Verification and initial annotation of NIA mouse 15K cDNA clone set

Unpublished (2001)

Other ESTs: H3094F01-5

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.

Plate: H3094 row: F column: 01

Seq primer: -21M13 Forward

High quality sequence stop: 358

POLYA=Yes.

Location/Qualifiers

1. .358

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="niaEST:H3094F01-3"

/db\_xref="taxon:10090"

/clone="H3094F01"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev\_stage="Clones arrayed from a variety of cDNA libraries"

/lab\_host="DH10B"

/clone\_lib="NIA Mouse 15K cDNA Clone Set"

/note="Vector: pSPORT1; Site 1: Sali; Site 2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5

embryos, and E12.5 female mesonephros/gonad) and one

newborn ovary cDNA library. Average insert size 1.5 kb.

All source libraries are cloned unidirectionally with

Oligo(dT)-Not primers. References include: (1)

Genome-wide expression profiling of mid-gestation

placenta and embryo using a 15,000 mouse developmental

cDNA microarray. 2000, Proc. Natl. Acad. Sci. U S A, 97:

9127-9132; (2) Large-scale cDNA analysis reveals phased

gene expression patterns during preimplantation mouse

development, 2000, Development, 127: 1737-1749; (3)

Genome-wide mapping of unselected transcripts from

extraembryonic tissue of 7.5-day mouse embryos reveals

chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN  
Query Match 80.8%; Score 20.2; DB 2; Length 358;  
Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAGAGCGCAGACTGCTCTCC 25  
|||||  
Db 168 AAAAAAGAGCGCAGACTGCTCTCC 144  
|||||

RESULT 12  
BH315230/c  
LOCUS  
DEFINITION BH315230 363 bp DNA linear GSS 03-DEC-2001  
CH230-159C13-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-159C13, genomic survey sequence.  
ACCESSION BH315230  
VERSION  
KEYWORDS  
SOURCE GSS.  
ORGANISM Rattus norvegicus (Norway rat)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Teegaye, G., Geer, K.,  
Riggs, F., de Jong, P., and Fraser, C.M.  
SHVARTSBEYN, A., GABREGORIS, E., OVERTON, L., RUSSELL, D., CHEN, D.,  
RAT BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)  
Other\_GSSs: CH230-159C13-TJ

TITLE  
JOURNAL  
COMMENT Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhac@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or\_ering\_information.htm). BAC end  
page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 159 row: C column: 13  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..363  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SnHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-159C13"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by  
Pieter de Jong"

ORIGIN  
Query Match 80.8%; Score 20.2; DB 9; Length 363;  
Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAGAGCGCAGACTGCTCTCC 25  
|||||  
Db 322 AAAAAAGAGCGCAGACTGCTCTCC 298  
|||||

RESULT 13  
BG083952  
LOCUS  
DEFINITION H3094F01-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
H3094F01 5', mRNA sequence.  
ACCESSION BG083952  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)

REFERENCE  
AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A.,  
Tanaka, T.S., Carter, M.G. and Ko, M.S.H.  
TITILE Verification and initial annotation of NIA mouse 15K cDNA clone set  
JOURNAL Unpublished (2001)  
COMMENT Other ESTs: H3094F01-3

Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Casell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.  
Plate: H3094 row: F column: 01  
Seq primer: -21M13 Reverse  
High quality sequence stop: 371  
POLYA=No.

FEATURES  
source Location/Qualifiers  
1..371  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:H3094F01-5"  
/db\_xref="taxon:10090"  
/clone="H3094F01"  
/sex="Clones arrayed from a variety of cDNA libraries"  
/dev\_stage="Clones arrayed from a variety of cDNA libraries"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse 15K cDNA Clone Set"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This  
clone is among a rearrayed set of 15,247 clones from 11  
embryo cDNA libraries (including preimplantation stage  
embryos from unfertilized egg to blastocyst, embryonic  
part of E7.5 embryos, extraembryonic part of E7.5  
embryos, and E12.5 female mesonephros/gonad) and one  
newborn ovary cDNA library. Average insert size 1.5 kb.  
All source libraries are cloned unidirectionally with  
Oligo(dT)-Not primers. References include: (1)  
Genome-wide expression profiling of mid-gestation  
placenta and embryo using a 15,000 mouse developmental  
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97:  
9127-9132; (2) Large-scale cDNA analysis reveals phase  
gene expression patterns during preimplantation mouse  
development, 2000, Development, 127: 1737-1749; (3)  
Genome-wide mapping of unselected transcripts from  
extraembryonic tissue of 7.5-day mouse embryos reveals  
enrichment in the t-complex and under-representation on  
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN  
Query Match 80.8%; Score 20.2; DB 2; Length 371;  
Best Local Similarity 88.0%; Pred. No. 1.6e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGCTCTCC 25  
|||||  
Db 180 AAAAAAGCGCAGACTGCTCTCC 204  
|||||



RESULT 14  
BH344887/c  
LOCUS  
DEFINITION CH230-183013.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-183013, genomic survey sequence.  
ACCESSION BH344887  
VERSION BH344887.1 GI:17275621  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM  
REFERENCE  
AUTHORS Zhao S., Shetty J., Shatsman S., Tesgaye G., Geer K.,  
Shivarsbeyn A., Gebregorgis E., Overton L., Russell D., Chen D.,  
Riggs F., de Jong P. and Fraser C.M.  
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: CH230-183013.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or ering information.htm). BAC end  
page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 183 row: O column: 13  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1..393  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SENHad/MCW"  
/db\_xref="taxon:10116"  
/clones="CH230-183013"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CHORI-230 Rat (BN/SENHad/MCW) BAC library produced by  
Pieter de Jong"

ORIGIN  
Query Match 80.8%; Score 20.2; DB 9; Length 393;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 332 AAAAAAGAGACCGACTGCTCTTCC 308  
|||||

RESULT 15  
CN669196  
LOCUS  
DEFINITION CN669196 400 bp mRNA linear EST 17-MAY-2004  
musculus cDNA clone NIA:A0876H02 IMAGE:30764533 5', mRNA sequence.  
ACCESSION CN669196  
VERSION CN669196.1 GI:47435647  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 400)  
Sharov, A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y.,  
VanBuren, V., Falco, G., Martin, P.R., Stagg, C.A., Basse, U.C.,  
Wang, Y., Carter, M.G., Hamatani, T., Aiba, K., Akutsu, H., Sharova, L.,  
Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S.,  
Nagaraja, R., Boheler, K.R., Raub, D., Hodes, R.J., Longo, D.L.,  
Schlessinger, D., Keller, J., Klotz, E., Klesoe, G., Umazawa, A.,  
Vescovi, A.L., Rossant, J., Kunath, T., Hogan, B.L., Curcio, A.,  
D'Urso, M., Kelso, J., Hide, W. and Ko, M.S.

TITLE  
JOURNAL  
COMMENT

Transcriptome analysis of mouse stem cells and early embryos  
PLOS Biol. 1 (3), 410-419 (2003)  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdnaelg@nigam.nih.gov  
Plate: A0876 row: H column: 02  
Seq primer: M13 Reverse  
High quality sequence stop: 400  
POLYA=No.

FEATURES  
source

1..400  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nia:EST:A0876H02-5"  
/db\_xref="taxon:10090"  
/clone="NIA:A0876H02 IMAGE:30764533"  
/tissue\_type="whole embryo including extraembryonic  
tissues at 13.5-days postcoitum"  
/dev\_stage="E13.5"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse E13.5 whole embryo cDNA library  
(Long)"  
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).  
This is a long-transcript enriched cDNA library (Ref.  
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total  
RNAs were extracted from 1 embryo at 13.5-days postcoitum.  
Double-stranded cDNAs were synthesized with an Oligo(dT)  
primer [Invitrogen]:  
5'-pgactagtcttagatcgccgcgcctttttttttt-3' from  
3ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to lone-linker LI-Sal4, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 3.0Kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 400;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 30 AAAAAAGAGCAGCTGCTCTTCC 54  
|||||

RESULT 16  
AW535711/c  
LOCUS  
DEFINITION AW535711 417 bp mRNA linear EST 06-MAR-2000  
UI-R-BS0-aod-b-06-0-UI.s1 UI-R-BS0 Rattus norvegicus cDNA clone  
UI-R-BS0-aod-b-06-0-UI 3', mRNA sequence.



ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AW535711 AW535711.1 EST. Rattus norvegicus (Norway rat) Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE AUTHORS TITLE JOURNAL COMMENT	1 (bases 1 to 417) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996) 8889548 Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at 13 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-24, >AT rich#Low complexity 216-364, >B2\$INE/B2 364-417, >BC1_MW#scrNA Seq primer: M13 Forward POLYA=yes.
FEATURES source	Location/Qualifiers 1..417 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague-Dawley" /db_xref="taxon:10116" /clone="UI-R-BS0-aod-b-06-0-UI" /dev_stage="embryonic 13 dpc" /lab_host="DH10B (Life Technologies)" /clone_lib="UI-R-BS0" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS0 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. TAG_TISSUE=embryo at 13 dpc TAG_LIB=UI-R-BS0 TAG_SEQ=AATCC"
ORIGIN	Query Match 80.8%; Score 20.2; DB 1; Length 417; Best Local Similarity 88.0%; Pred. No. 1.7e+02; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  Qy 1 AAAAAAGAGCGAGACTGGTCTTCC 25       Db 368 AAAAAAGAGCGAGCTGCTCTTCC 344 
RESULT 17 BZ207783 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	BZ207783 420 bp DNA linear GSS 11-OCT-2002 CH230-475011.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-475011, genomic survey sequence. BZ207783 BZ207783.1 GI:23865835 GSS. Rattus norvegicus (Norway rat) Rattus norvegicus
REFERENCE AUTHORS TITLE JOURNAL COMMENT	1 (bases 1 to 420) Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M. Rat BAC End Sequences from Library CHORI-230 MboI segment Unpublished (1999) Other_GSSs: CH230-475011.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.html). BAC end page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html Plate: 475 row: 0 column: 11 Seq primer: T7 Class: BAC ends. Location/Qualifiers 1..420 /organism="Rattus norvegicus" /mol_type="genomic DNA" /strain="BN/SSNHsd/MCW" /db_xref="taxon:10116" /clone="CH230-475011" /sex="Female" /cell_type="Brain" /clone_lib="CHORI-230 Segment 2" /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN	Query Match 80.8%; Score 20.2; DB 9; Length 420; Best Local Similarity 88.0%; Pred. No. 1.7e+02; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  Qy 1 AAAAAAGAGCGAGACTGGTCTTCC 25       Db 111 AAAAAAGAGCGACTGCTCTTCC 135 
RESULT 18 BY567189/C LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	BY567189 430 bp mRNA linear EST 15-DEC-2002 BY567189 RIKEN full-length enriched, activated spleen Mus musculus cDNA clone F830203P15 3', mRNA sequence. BY567189 BY567189.1 GI:26902371 EST. Mus musculus (house mouse) Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430) Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Balgarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
REFERENCE AUTHORS	1 (bases 1 to 430) Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Balgarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,

Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagahima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

**TITLE**  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

**JOURNAL PUBLISHED**  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

**COMMENT**  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

**TITLE**  
RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

**TITLE**  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

**TITLE**  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

**TITLE**  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

**TITLE**  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
source  
Location/Qualifiers  
1..430  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="taxon:10090"  
/clone="F830203P15"  
/tissue type="activated spleen"  
/clone\_lib="RIKEN full-length enriched, activated spleen"

**ORIGIN**  
Query Match 80.8%; Score 20.2; DB 5; Length 430;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY**  
1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
||||| ||| ||||| ||||| |||||  
76 AAAAAAGAGCGCAGACTGGTCTTCC 52

**Db**

**RESULT 19**  
BP417531/c  
LOCUS  
DEFINITION  
BP417531 Homo sapiens small intestine Homo sapiens cDNA clone  
HIE00567r 3', mRNA sequence.

**ACCESSION**  
BP417531  
VERSION  
BP417531.1 GI:66783784

**KEYWORDS**  
EST.  
Homo sapiens (human)  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Primates; Catarrhini;  
Hominidae; Homo.

**REFERENCE**  
1 (bases 1 to 443)  
Takeda, J., Jin, L. and Horikawa, Y.  
Expression profile of mRNAs from human pancreatic islet tumors  
Unpublished (2005)

**JOURNAL COMMENT**  
Contact: Yukio Horikawa  
Laboratory of Molecular Genetics  
Institute for Molecular and Cellular Regulation, Gunma University  
Showa-machi 3-39-15, Maebashi, Gunma, 371-8512, Japan  
Tel: 81-27-220-8832  
Fax: 81-27-220-8889  
Email: yhorikawa@showa.gunma-u.ac.jp.

**FEATURES**  
source  
Location/Qualifiers  
1..443  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HIE00567r"  
/tissue type="small intestines"  
/clone\_lib="Homo sapiens small intestine"

**ORIGIN**  
Query Match 80.8%; Score 20.2; DB 3; Length 443;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY**  
1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
||||| ||| ||||| ||||| |||||  
76 AAAAAAGAGCGCAGACTGGTCTTCC 52

**Db**

**RESULT 19**  
BP417531/c  
LOCUS  
DEFINITION  
BP417531 Homo sapiens small intestine Homo sapiens cDNA clone  
HIE00567r 3', mRNA sequence.

**ACCESSION**  
BP417531  
VERSION  
BP417531.1 GI:66783784

**KEYWORDS**  
EST.  
Homo sapiens (human)  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Primates; Catarrhini;  
Hominidae; Homo.

**REFERENCE**  
1 (bases 1 to 443)  
Takeda, J., Jin, L. and Horikawa, Y.  
Expression profile of mRNAs from human pancreatic islet tumors  
Unpublished (2005)

**JOURNAL COMMENT**  
Contact: Yukio Horikawa  
Laboratory of Molecular Genetics  
Institute for Molecular and Cellular Regulation, Gunma University  
Showa-machi 3-39-15, Maebashi, Gunma, 371-8512, Japan  
Tel: 81-27-220-8832  
Fax: 81-27-220-8889  
Email: yhorikawa@showa.gunma-u.ac.jp.

**FEATURES**  
source  
Location/Qualifiers  
1..443  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HIE00567r"  
/tissue type="small intestines"  
/clone\_lib="Homo sapiens small intestine"

**ORIGIN**  
Query Match 80.8%; Score 20.2; DB 3; Length 443;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY**  
1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
||||| ||| ||||| ||||| |||||  
180 AAAAAAGAGCTCAACGGGTCTTCC 156

**Db**

**RESULT 20**  
AZ960237/c  
LOCUS  
DEFINITION  
AZ960237 Mouse 10kb plasmid tUGC2M library Mus musculus genomic  
clone UGC2M0228P09 F, genomic survey sequence.

**ACCESSION**  
AZ960237  
VERSION  
AZ960237.1 GI:13831464

**KEYWORDS**  
GSS.  
Mus musculus (house mouse)

**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

**REFERENCE**  
1 (bases 1 to 464)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von  
Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

**JOURNAL COMMENT**  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

**QY**  
1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
||||| ||| ||||| ||||| |||||  
180 AAAAAAGAGCTCAACGGGTCTTCC 156

**Db**

**RESULT 20**  
AZ960237/c  
LOCUS  
DEFINITION  
AZ960237 Mouse 10kb plasmid tUGC2M library Mus musculus genomic  
clone UGC2M0228P09 F, genomic survey sequence.

**ACCESSION**  
AZ960237  
VERSION  
AZ960237.1 GI:13831464

**KEYWORDS**  
GSS.  
Mus musculus (house mouse)

**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarctontoglires; Glires; Rodentia;  
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

**REFERENCE**  
1 (bases 1 to 464)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von  
Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

**JOURNAL COMMENT**  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0228 row: P column: 09  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 464.  
 Location/Qualifiers  
 1. .464  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0228P09"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC2M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G1[4732114|9b|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## FEATURES

source

1. .464  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0228P09"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC2M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G1[4732114|9b|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 464;  
 Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25

Db 413 AAAAAAGCGCAGCTGCTCTTCC 389

## RESULT 21

BH260028/c

LOCUS

DEFINITION CH230-113A8-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-113A8, genomic survey sequence.

ACCESSION BH260028

VERSION BH260028.1

KEYWORDS GI:17163956

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;  
 1 (bases 1 to 480)

## REFERENCE

AUTHORS

Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K., Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.  
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment Unpublished (1999)

TITLE

JOURNAL

COMMENT

Other GSSs: CH230-113A8.TJD  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@cigr.org  
 Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering information.htm). BAC end page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 113 row: A column: 8  
 Seq primer: T7  
 Class: BAC ends.

Location/Qualifiers

## FEATURES

source

1. 480  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-113A8"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 1"  
 /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 480;  
 Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25

Db 174 AAAAAAGCGCAGCTGCTCTTCC 150

## RESULT 22

BY569957/c

LOCUS

DEFINITION BY569957 RIKEN full-length enriched, activated spleen Mus musculus  
 CDNA clone F830216K22 3', mRNA sequence.

ACCESSION BY569957

VERSION BY569957.1

KEYWORDS GI:26905139

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
 1 (bases 1 to 481)

## REFERENCE

AUTHORS

Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaoka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kurohchi,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., Mckenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,X., Watanabe,Y., Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,

Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,  
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,B., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
12466851

## TITLE

## JOURNAL

## PUBMED

## COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics  
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
Trust/MRC building Addenbrookes Hospital Cambridge) whose  
assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

## FEATURES

## source

Location/Qualifiers  
1. .481  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="taxon:10090"  
/clones="F030216K22"  
/tissue\_types="activated spleen"  
/clone\_lib="RIKEN full-length enriched, activated spleen"

## ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 481;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25

Db 123 AAAAAAGAGCTGCTGCTTCC 99

RESULT 23

Bp775662/c

## LOCUS

BP775662 mouse (C57BL/6) pancreatic islet library with  
recombination-based method Mus musculus cDNA clone mif25067 3',  
mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

BP775662

BP775662.1 GI:50238179

EST.

## SOURCE

## ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 482)

Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H.,

Takeda,J., Ohara,O. and Seino,S.

Construction of a multi-functional cDNA library specific for mouse  
pancreatic islets and its application to microarray

DNA Res 11 (5), 315-323 (2004)

15747579

Contact: Susumu Seino

Division of Cellular and Molecular Medicine

Kobe University Graduate School of Medicine

7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan

Tel: 81-78-382-5360

Fax: 81-78-382-5370

Email: [seino@med.kobe-u.ac.jp](mailto:seino@med.kobe-u.ac.jp).

Location/Qualifiers

## FEATURES

## source

1. .482

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="mif25067"

/sex="male"

/tissue\_type="pancreatic islet"

/dev\_stage="adult"

/clone\_lib="mouse (C57BL/6) pancreatic islet library with

recombination-based method"

recombination-based method"

## ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 482;

Best Local Similarity 88.0%; Pred. No. 1.7e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25

Db 363 AAAAAAGAGCTGCTGCTTCC 339

RESULT 24

BI301386/c

## LOCUS

BI301386 506 bp mRNA linear EST 20-JUL-2001  
UI-R-DL0-cio-f-02-0-UI.sl UI-R-DL0 Rattus norvegicus cDNA clone  
UI-R-DL0-cio-f-02-0-UI 3', mRNA sequence.

BI301386

BI301386.1 GI:14977666

EST.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

1 (bases 1 to 506)

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

non-normalized rat tongue library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-155, >B2#SINE/B2 164-255, >ID RN#SINE/ID 258-294, >B2#SINE/B2 299-493, >B2#SINE/B2 Seq primer: M13 Forward POLYA=Yes.

#### FEATURES

source  
Location/Qualifiers  
1. 506  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DL0-cio-f-02-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="PH10B (Life Technologies)"  
/clone\_lib="UI-R-DL0"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-DL0 library is a non-normalized Rat Tongue library constructed in pT7T3 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag GCGAA between the Not I cloning site and dT18 stretch.  
TAG\_TISSUE=rat tongue  
TAG\_LIB=UI-R-DL0  
TAG\_SEQ=GCGAA"

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 506;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
|||||  
Db 159 AAAAAAGAGCAGTCTGCTCTTCC 135  
|||||

RESULT 25  
AZ875623/c  
LOCUS  
DEFINITION  
2M0190D13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0190D13 F, genomic survey sequence.

ACCESSION  
AZ875623  
VERSION  
AZ875623.1 GI:13085816  
KEYWORDS  
GSS.

SOURCE  
Mus musculus (house mouse)

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

#### REFERENCE

AUTHORS  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islan,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D.,Weiss,R.

TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

#### JOURNAL

COMMENT  
Unpublished (2000)

Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0190 row: D column: 13

Seq primer: CGTTGTAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 509.

#### FEATURES

source

#### Location/Qualifiers

1. 509  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0190D13"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 509;  
Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
|||||  
Db 356 AAAAAAGAGCTCTGACTGCTCTTCC 332  
|||||

#### RESULT 26

BB756913/c

LOCUS

DEFINITION

BB756913 RIKEN full-length enriched, melanocyte Mus musculus cDNA  
clone G270079U17 3', mRNA sequence.

ACCESSION

BB756913

VERSION

BB756913.1 GI:16203240

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 528)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,

Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,

Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,

Matuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,

Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,

Tanaka,T., Tonaru,A., Toya,T., Watahiki,A., Yasunishi,A.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
 Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.  
 and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

**FEATURES**  
 source  
 1. 528  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G270079J17"  
 /cell\_type="melanocyte"  
 /clone\_lib="RIKEN full-length enriched, melanocyte"

**ORIGIN**  
 Query Match 80.8%; Score 20.2; DB 2; Length 528;  
 Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**Qy** 1 AAAAAAGAGCGACACTGGTCTTCC 25  
 |||||  
**Db** 383 AAAAAAGAGCGACACTGGTCTTCC 359  
 |||||

**RESULT 27**  
**BH349233/c**  
**LOCUS** BH349233 536 bp DNA linear GSS 03-DEC-2001  
**DEFINITION** CH230-81B24.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-81B24, genomic survey sequence.  
**ACCESSION** BH349233  
**VERSION** BH349233.1 GI:17279967  
**KEYWORDS** GSS.  
**SOURCE** Rattus norvegicus (Norway rat)  
**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Rattus.  
**REFERENCE** 1 (bases 1 to 536)  
**AUTHORS** Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,  
 Shwartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D.,  
 Riggs, P., de Jong, P. and Fraser, C.M.  
**TITLE** Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
**JOURNAL** Unpublished (1999)  
**COMMENT** Other GSSs: CH230-81B24.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.choi.org).  
 Clones may be purchased from BACPAC Resources  
 ([http://www.chori.org/bacpac/orering\\_information.htm](http://www.chori.org/bacpac/orering_information.htm)). BAC end

page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
 Plate: 81 row: B column: 24  
 Seq primer: SP6  
 Class: BAC ends.

**FEATURES**  
 source  
 1. 536  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-81B24"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 1"  
 /notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

**ORIGIN**  
 Query Match 80.8%; Score 20.2; DB 9; Length 536;  
 Best Local Similarity 88.0%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**Qy** 1 AAAAAAGAGCGACACTGGTCTTCC 25  
 |||||  
**Db** 170 AAAAAAGAGCGACACTGGTCTTCC 146  
 |||||

**RESULT 28**  
**BG088328**  
**LOCUS** H3151G04 574 bp mRNA linear EST 18-DEC-2003  
**DEFINITION** H3151G04-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
 H3151G04 5', mRNA sequence.  
**ACCESSION** BG088328  
**VERSION** BG088328.2 GI:40109258  
**KEYWORDS** EST.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 574)  
**AUTHORS** Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X.,  
 Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H.,  
 Wood, W.H. III, Becker, K.G. and Ko, M.S.H.  
**TITLE** Genome-wide expression profiling of mid-gestation placenta and  
 embryo using a 15,000 mouse developmental cDNA microarray  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
**COMMENT** 10922068  
 On Jan 26, 2001 this sequence version replaced gi:12570892.  
 Other ESTs: H3151G04-3  
 Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgaun.grc.nia.nih.gov  
 This clone set has been freely distributed to the community. Please  
 visit <http://lgaun.grc.nia.nih.gov/cdna/15k.html> for details.  
 Plate: H3151 row: G column: 04  
 Seq primer: -21M13 Reverse  
 High quality sequence stop: 574  
 POLYA-No.

**FEATURES**  
 source  
 1. 574  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="niaEST:H3151G04-5"  
 /db\_xref="taxon:10090"  
 /clone="H3151G04"  
 /sex="Clones arrayed from a variety of cDNA libraries"  
 /dev\_stage="Clones arrayed from a variety of cDNA  
 libraries"



/lab\_host="DH10B"  
 /clone\_lib="NIA Mouse 15K cDNA Clone Set"  
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. 2000. Proc. Natl. Acad. Sci. U S A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 574;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGCTCTTCC 25  
 |||||  
 Db 256 AAAAAAGAGCACATCTGCTCTTCC 280

## RESULT 29

AZ288000 577 bp DNA linear GSS 27-JUL-2000  
 LOCUS RPCI-23-142E9.TV RPCI-23 Mus musculus genomic clone RPCI-23-142E9,  
 DEFINITION genomic survey sequence.

ACCESSION AZ288000  
 VERSION AZ288000.1 GI:9529786  
 KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

## REFERENCE

Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de  
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S.,  
 1 (bases 1 to 577)  
 Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSSs: RPCI-23-142E9.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac ends/mouse/bac\_end\_intro.html

Plate: 142 row: E column: 9

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 577

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

## FEATURES

## source

/db\_xref="taxon:10090"  
 /clone="RPCI-23-142E9"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /clone\_lib="RPCI-23"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 577;

Best Local Similarity 88.0%; Pred. No. 1.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGACACTGCTCTTCC 25

|||||

Db 455 AAAAAAGAGTGTGACTGATCTTCC 479

## RESULT 30

BH261566

LOCUS

DEFINITION

CH230-90J19.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

BH261566

VERSION

BH261566.1 GI:17167188

KEYWORDS

GSS.

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 586)

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,

Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,

Riggs, F., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other GSSs: CH230-90J19.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or ering information.html). BAC end

page: http://www.tigr.org/tdb/bac ends/rat/bac\_end\_intro.html

Plate: 90 row: J column: 19

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 586

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/strain="BN/SSNHsd/MCW"

/db\_xref="taxon:10116"

/clone="CH230-90J19"

/sex="Female"

/cell\_type="Brain"

/clone\_lib="CHORI-230 Segment 1"

/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

## ORIGIN



Query Match 80.8%; Score 20.2; DB 9; Length 586;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTTCC 25  
 |||||  
 Db 93 AAAAAAGAGCGCAGCTGCTCTTCC 117

RESULT 31  
 AZ357207/c  
 LOCUS  
 DEFINITION  
 1M0098D08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0098D08 R, genomic survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von  
 Niederhausen, A., and Wright, D., Weis, R.

TITLE  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0098 row: D column: 08  
 Seq primer: CACACAGGAAACAGCTATGACC  
 Class: plasmid ends

High quality sequence stop: 607.

Location/Qualifiers

FEATURES  
 source  
 1. 607  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0098D08"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 [gi|4732114|gb|AF129072.1], a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

# ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 607;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTTCC 25  
 |||||  
 Db 289 AAAAAAGAGCGCAGCTGCTCTTCC 265

RESULT 32  
 CN542163/c

LOCUS  
 DEFINITION  
 CN542163  
 UI-R-EA0-ckx-d-03-0-UI.s1 UI-R-EA0 Rattus norvegicus cDNA clone  
 UI-R-EA0-ckx-d-03-0-UI 3', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE  
 AUTHORS  
 TITLE  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery

Bonaldi, M.F., Lennon, G., and Soares, M.B.  
 Genome Res. 6 (9), 791-806 (1996)

JOURNAL  
 COMMENT  
 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu  
 Tissue Procurement: Tom Freeman, Sanger Center  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/rat.html

The following repetitive elements were found  
 in this cDNA  
 sequence: 1-171, >L1PB2#LINE/L1 (matched complement) 165-281,  
 >BC1 MM#scrna (matched complement) 282-321, >L1PB3#LINE/L1 (matched  
 complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source  
 1. 625  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-EA0-ckx-d-03-0-UI"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-EA0"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-EA0 is a  
 non-normalized Rat ileum library (RIL) constructed in  
 pT73 PAC vector according to the procedure described by  
 Bonaldi, Lennon & Soares (Genome Research Genome 6:  
 791-806, 1996). The oligonucleotide used to prime first  
 strand synthesis contained the sequence tag CGACACTCG  
 between the Not I cloning site and dT18 stretch. The Rat  
 ileum tissue was provided by Tom Freeman of the Sanger  
 Center.  
 TAG TISSUE=ileum  
 TAG LIB=UI-R-EA0  
 TAG\_SEQ=CGACACTCG"

# ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 625;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGGCTCTTCC 25

Db 175 AAAAAAGAGACTGACTGCTCTTCC 151

RESULT 33  
 BZ116863  
 LOCUS  
 DEFINITION CH230-51006.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 CH230-51006, genomic survey sequence.

ACCESSION BZ116863

VERSION BZ116863

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 632)  
 Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Gear,K.,  
 Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
 Riggs,F., de Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 MbOI segment

Unpublished (1999)

Other GSSs: CH230-51006.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering information.html). BAC end

page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html

Plate: 510 row: 0 column: 6

Seq primer: T7

Class: BAC ends.

FEATURES

source

1. .632  
 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-51006"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /notes="Vector: pTARBAC1.3; Site 1: MbOI; Site 2: MbOI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 632;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGGCTCTTCC 25

Db 418 AAAAAAGAGCGCTGACTGCTCTTCC 442

RESULT 34  
 AG282644/c  
 LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 702;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGGCTCTTCC 25

Db 572 AAAAAAGAGACTGACTGCTCTTCC 548

RESULT 35

BZ196833/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus molossinus DNA, clone:MSMg01-054M22.TJ, genomic survey  
 sequence.

AG282644

AG282644.1 GI:47855521

GSS.

Mus musculus molossinus (Japanese wild mouse)

Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

1

Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,  
 Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and  
 Shiroishi,T.

Contribution of Asian mouse subspecies Mus musculus molossinus to  
 genomic constitution of strain C57BL/6J, as defined by BAC-end

sequence-SNP analysis

Genome Res. 14 (12), 2439-2447 (2004)

15574823

2 (bases 1 to 702)

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/.

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kunia Abe (abe@rtc.riken.jp).

Tsukuba Institute, Bio Resource Center,

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI.

R.Site 2 : EcoRI.

Location/Qualifiers

1. .702

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-054M22.TJ"

/sex="male"

/tissue\_type="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

BZ196833  
 CH230-269H12.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 CH230-269H12, genomic survey sequence.

BZ196833

BZ196833.1 GI:23854885

GSS.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciuognathi; Muroidea; Muridae; Murinae; Rattus.  
 1 (bases 1 to 708)  
**REFERENCE**  
**AUTHORS** Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K., Shvartsbeyn,A., Gebregioris,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.  
**TITLE** Rat BAC End Sequences from Library CHORI-230 MboI segment  
**JOURNAL** Unpublished (1999)  
**COMMENT** Other GSSs: CH230-269H12.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
 availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
 Clones may be purchased from BACPAC Resources  
 (<http://www.chori.org/bacpac/oreringinformation.htm>). BAC end  
 page: [http://www.tigr.org/tldb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html)  
 Plate: 269 row: H column: 12  
 Seq primer: T7  
 Class: BAC ends.

#### FEATURES

source

1. 708  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SsNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-269H12"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: pTARAC1.3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by  
 Pieter de Jong"

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 708;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 AAAAAAGCGCGCTGACTGCTTCC 25  
 |||||  
 Db 584 AAAAAAGCGCGCTGACTGCTTCC 560

RESULT 36  
 AG562735/c  
 LOCUS AG562735.1 GI:48323433  
 DEFINITION Mus musculus molossinus DNA, clone:MSMg01-482J18.TJ, genomic survey  
 sequence.  
 ACCESSION AG562735  
 VERSION AG562735.1  
 KEYWORDS GSS.  
 SOURCE Mus musculus molossinus (Japanese wild mouse)  
 ORGANISM Mus musculus molossinus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciuognathi; Muridae; Murinae; Mus.

REFERENCE 1  
 AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,  
 Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and  
 Shiroishi,T.  
 TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to  
 genomic constitution of strain C57BL/6J, as defined by BAC-end  
 sequence-SNP analysis  
 JOURNAL Genome Res. 14 (12), 2439-2447 (2004)  
 PUBMED 15574823  
 REFERENCE 2 (bases 1 to 714)  
 AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
 TITLE Direct Submission

#### JOURNAL

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
 (E-mail:hattori@gsc.riken.jp, URL:<http://hgp.gsc.riken.go.jp/>,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the mouse BAC library MSMg01. For BAC  
 library availability, please contact Kunya Abe ([abe@rtc.riken.jp](mailto:abe@rtc.riken.jp)).  
 Tsukuba Institute, Bio Resource Center,  
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
 Koyadai, Tsukuba, 305-0074 Japan  
 phone: 81-298-36-9189, fax: 81-298-36-9199  
 e-mail: [abe@rtc.riken.jp](mailto:abe@rtc.riken.jp)

#### PRIMERS

Sequencing : TJ

#### LIBRARY

Vector : pBACe3.6

R.Site 1 : BcoRI

R.Site 2 : EcoRI

#### FEATURES

source

1. 714  
 /organism="Mus musculus molossinus"  
 /mol\_type="genomic DNA"  
 /sub\_species="molossinus"  
 /db\_xref="taxon:57486"  
 /clone="MSMg01-482J18.TJ"  
 /sex="male"  
 /tissue type="mixture of kidney and spleen"  
 /clone\_lib="MSMg01-Mouse Male BAC Library"

#### ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 714;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 AAAAAAGCGCGCTGCTTCC 25  
 |||||  
 Db 519 AAAAAAGCGCGCTGCTTCC 495

#### RESULT 37

BG298680/c

LOCUS BG298680.1

DEFINITION 602396890F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:4511566 5',

mRNA sequence.

ACCESSION BG298680

VERSION BG298680.1

KEYWORDS GI:13063576

SOURCE EST.

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciuognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 726)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10395 row: c column: 23

High quality sequence stop: 726.

Location/Qualifiers

1. 726

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:4511566"

/tissue type="retina"  
 /lab host="DH10B (phage-resistant)"  
 /clone lib="NIH\_MGC\_94"  
 /note="Organ: eye; Vector: PCMV-SPORT6; Site: 1: NotI;  
 Site: 2: SalI; Cloned unidirectionally; oligo-dt primed.  
 Average insert size 3.3 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 726;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
 |||||  
 Db 167 AAAAAAGCAGCTGCTCTTCC 143

RESULT 38  
 AZ430769/c  
 LOCUS  
 DEFINITION 736 bp DNA linear GSS 03-OCT-2000  
 1M0215G22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0215G22 F, genomic survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 736)

REFERENCE  
 AUTHORS  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
 Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von  
 Niederhausern, A. and Wright, D. Weis, R.

TITLE  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00

Plate: 0215 row: G column: 22

Seq primer: CGTGTAAACGACGGCCAGT

Class: plasmid ends

High quality sequence stop: 736.

Location/Qualifiers

1..736

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0215G22"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptored DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid RL. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptored mouse DNA was annealed to  
 adaptored vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 736;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
 |||||  
 Db 343 AAAAAAGCAGCTGCTCTTCC 319

RESULT 39  
 BZ173903  
 LOCUS  
 DEFINITION 779 bp DNA linear GSS 11-OCT-2002  
 CH230-493G21.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 CH230-493G21, genomic survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.  
 1 (bases 1 to 779)

REFERENCE  
 AUTHORS  
 Zhao, S., Shetty, J., Shatsman, S., Tseng, G., Geer, K.,  
 Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D.,  
 Riggs, F., de Jong, P. and Fraser, C.M.  
 Rat BAC End Sequences from Library CHORI-230 MboI segment  
 Unpublished (1999)

TITLE  
 JOURNAL  
 COMMENT  
 Other\_GSSs: CH230-493G21.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering.information.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 493 row: G column: 21  
 Seq primer: SP6  
 Class: BAC ends

Location/Qualifiers  
 1..779  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-493G21"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
 Pieter de Jong"

FEATURES  
source

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 779;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGCTCTTCC 25  
 |||||  
 Db 324 AAAAAAGCGCAGACTGCTCTTCC 348  
 |||||

RESULT 40  
 BH292979  
 LOCUS CH230-92B14.TJB CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 DEFINITION CH230-92B14, genomic survey sequence.

ACCESSION BH292979  
 VERSION BH292979.1 GI:17205387

KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 789)  
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K.,  
 Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
 Riggs,F., de Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: CH230-92B14.TVB

Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering\_information.htm). BAC end  
 page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 92 row: B column: 14  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..789  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SENHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-92B14"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 1"  
 /note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
 Pieter de Jong"

Query Match 80.8%; Score 20.2; DB 9; Length 789;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGCTCTTCC 25  
 |||||  
 Db 590 AAAAAAGCGCAGACTGCTCTTCC 614  
 |||||

RESULT 41  
 BZ277088  
 LOCUS CH230-308G22.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 DEFINITION CH230-308G22, genomic survey sequence.

ACCESSION BZ277088

VERSION BZ277088.1 GI:24000638  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 809)  
 AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K.,  
 Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
 Riggs,F., de Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: CH230-308G22.TV

Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering\_information.htm). BAC end  
 page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 308 row: G column: 22  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..809  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SENHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-308G22"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
 Pieter de Jong"

Query Match 80.8%; Score 20.2; DB 9; Length 809;  
 Best Local Similarity 88.0%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGCTCTTCC 25  
 |||||  
 Db 583 AAAAAAGCGCAGACTGCTCTTCC 607  
 |||||

RESULT 42  
 AG295893/c  
 LOCUS AG295893  
 DEFINITION Mus musculus molossinus (Japanese wild mouse) sequence.

ACCESSION AG295893  
 VERSION AG295893.1 GI:47868847

KEYWORDS Mus musculus molossinus (Japanese wild mouse)  
 SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1  
 AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,  
 Szawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaiki,K. and  
 Shiroishi,T.

TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to  
 genomic constitution of strain C57BL/6J, as defined by BAC-end

```

sequence-SNP analysis
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 816)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Maehira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center, Research (RIKEN) 3-1-1
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
Location/Qualifiers
1..816
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-072F03.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 80.8%; Score 20.2; DB 10; Length 816;
Best Local Similarity 88.0%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
|||||
Db 351 AAAAAAGCGCAGACTGGTGGCCC 327

RESULT 43
BQ959424/c
LOCUS
DEFINITION
AGENCOURT_8951799 NCI CGAP Co24 Mus musculus cDNA clone
IMAGE:6476660 5', mRNA sequence.
BQ959424
BQ959424.1 GI:22374902
EST.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 831)
NITH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14017 Row: f Column: 21
High quality sequence stop: 394.

sequence-SNP analysis
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 816)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Maehira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center, Research (RIKEN) 3-1-1
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
Location/Qualifiers
1..831
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6476660"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/notes="Organ: Colon; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

ORIGIN
Query Match 80.8%; Score 20.2; DB 5; Length 831;
Best Local Similarity 88.0%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
|||||
Db 185 AAAAAAGAGCGACTGACTGCTCTTCC 161

RESULT 44
B275313
LOCUS
DEFINITION
CH230-450K22.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-450K22, genomic survey sequence.
ACCESSION
B275313
VERSION
B275313.1 GI:23997291
KEYWORDS
GSS.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 1052)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-450K22.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 450 Row: K Column: 22
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..1052
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-450K22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/notes="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong"

```

## ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 1052;  
 Best Local Similarity 88.0%; Pred. No. 1.9e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25  
 |||||  
 DB 715 AAAAAAGAGCGCAGCTGCTCTTCC 739

## RESULT 45

AA138597 401 bp mRNA linear EST 11-FEB-1997  
 LOCUS mc92c11.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
 DEFINITION IMAGE:S86196 5', mRNA sequence.

ACCESSION AA138597  
 VERSION AA138597.1 GI:1700809  
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 401)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:360844  
 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 389.

## FEATURES

source  
 1..401  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="NIH Swiss"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:S86196"  
 /sex="pooled"  
 /tissue\_type="heart"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /note="Organ: heart; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'  
 adaptor sequence: 5' GAATTCGACGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

## ORIGIN

Query Match 79.2%; Score 19.8; DB 1; Length 401;  
 Best Local Similarity 91.3%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTT 23  
 |||||  
 DB 379 AAAAAAGAGCGCGCAGCTGCTCTT 401

## RESULT 46

CO034709  
 LOCUS

DEFINITION kb Coccidioides posadasii cDNA clone C1FBV39 5' end, mRNA sequence.

ACCESSION CO034709

VERSION CO034709.1 GI:48572066

KEYWORDS EST.

SOURCE Coccidioides posadasii

ORGANISM Coccidioides posadasii

REFERENCE 1 (bases 1 to 868)

AUTHORS Gardner,M.J. and Cole,G.T.

TITLE Analysis of gene expression in Coccidioides posadasii mycelia and

JOURNAL spherules via expressed sequence tags

COMMENT Unpublished (2003)

Other\_ESTs: EST813092

Contact: Gardner MJ

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301 838 3519

Fax: 301 838 0208

Email: gardner@tigr.org

Seq primer: M13 Reverse.

Location/Qualifiers

1..868

/organism="Coccidioides posadasii"

/mol\_type="mRNA"

/strain="C735"

/db\_xref="taxon:199306"

/clone="C1FBV39"

/dev\_stage="spherules"

/lab\_host="E. coli DH10B, T1 phage resistant"

/clone\_lib="Coccidioides posadasii spherule cDNA library,  
 0.5 to 5.3 kb"

/notes="Vector: pExpress 1; Site\_1: Not 1; Site\_2: Eco RV;  
 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
 kb"

Query Match 79.2%; Score 19.8; DB 7; Length 868;  
 Best Local Similarity 91.3%; Pred. No. 2.9e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTT 23  
 |||||

DB 24 AAAAAAGAGCGCGACTGGTCTT 46

## RESULT 47

AA590658

LOCUS

DEFINITION

AA590658

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT



Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:563454

Putative full length read  
vector to vector length is 160.

#### FEATURES

Location/Qualifiers

```
1..159
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/db_xref="taxon:10090"
/clone="IMAGE:991174"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/clone_lib="Knowles Solter mouse blastocyst B1"
/note="Organ: embryo; Vector: pSPORT; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dtr):
5'-CGGTCCAGCTGACCGTGTGTTTTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
```

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 159;  
Best Local Similarity 95.2%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
|||||  
DB 14 AAGAGCGCAGACTGGTCTTCC 34

#### RESULT 48

AA823597 277 bp mRNA linear EST 17-FEB-1998  
LOCUS vw38a05.r1 Soares mammary\_gland NbMMG Mus musculus cDNA clone  
DEFINITION IMAGE:1246064 5', mRNA sequence.

AA823597

AA823597 GI:2893465

EST.

Mus musculus (house mouse)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 277)

REFERENCE  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:659752

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 241.

Location/Qualifiers

#### FEATURES

source

```
1..277
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1246064"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary_gland NbMMG"
/note="Organ: mammary gland; Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dtr) primer [5',  
TGTTACCAATCTGAAGTGGAGCGCGCGAATGCTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."
```

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 277;  
Best Local Similarity 95.2%; Pred. No. 3.8e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25

|||||

DB 243 AAGAGCGCAGACTGGTCTTCC 263

#### RESULT 49

CR109101/c

LOCUS

DEFINITION

Reverse strand read from insert in 5'HPRT insertion targeting and  
chromosome engineering clone MHPN151g12, genomic survey sequence.

CR109101

CR109101.1 GI:49856516

GSS; genome survey sequence; MICER.

Mus musculus (house mouse)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 292)

REFERENCE  
AUTHORS Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L.,  
Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y.,  
Rogers, J. and Bradley, A.

Direct Submission

Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. http://www.sanger.ac.uk/MICER

Location/Qualifiers

source

1..292

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="MHPN151g12"

/clone\_lib="MHPN"

#### ORIGIN

Query Match 77.6%; Score 19.4; DB 1; Length 292;  
Best Local Similarity 95.2%; Pred. No. 3.8e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25

|||||

DB 185 AAGAGCGCAGACTGGTCTTCC 165

LOCUS BE197804 343 bp mRNA linear EST 26-JUN-2000  
DEFINITION us77e02.x1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:3257210 3',  
mRNA sequence.  
ACCESSION BE197804  
VERSION BE197804.1 GI:8709907  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 343)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

FEATURES  
source MGI:1068574  
Seq primer: -40UP from Gibco  
High quality sequence stop: 342.  
Location/Qualifiers  
1..343  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3257210"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam4"  
/note="Organ: Mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

ORIGIN  
Query Match 77.6%; Score 19.4; DB 2; Length 343;  
Best Local Similarity 95.2%; Pred. No. 3.9e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 5 AAGAGCGCAGACTGGTCTTCC 25  
|||||  
Db 167 AAGAGCGCAGACTGGTCTTCC 147

Search completed: February 3, 2006, 22:02:48  
Job time : 2960.67 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 80.6667 Seconds  
(without alignments)  
550.897 Million cell updates/sec

Title: US-10-719-900-35

Perfect score: 25

Sequence: 1 aaaaagagcgactgtcttcc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6 COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfileseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.8	71.2	11176	3	US-09-465-978A-51
2	17.6	70.4	601	3	US-09-949-016-13905
3	17.6	70.4	1285	3	US-09-999-833A-528
4	17.6	70.4	1285	3	US-10-020-445A-528
5	17.6	70.4	2602	3	US-09-999-833A-525
6	17.6	70.4	2602	3	US-10-020-445A-525
7	17.6	70.4	2652	3	US-09-667-373-3
8	17.6	70.4	2652	3	US-10-305-413-3
9	17.6	70.4	3291	2	US-08-574-763-1
10	17.6	70.4	3377	3	US-09-819-989-1
11	17.6	70.4	3377	3	US-10-273-922-1
12	17.6	70.4	3377	3	US-10-681-222-1
13	17.6	70.4	19650	3	US-09-819-989-3
14	17.6	70.4	19650	3	US-10-273-922-3
15	17.6	70.4	19650	3	US-10-681-222-3
16	17.6	70.4	155266	3	US-09-949-016-13870
17	17.6	68.8	601	3	US-09-949-016-170047
18	17.2	68.8	13743	3	US-09-949-016-16538
19	17.6	68.0	501	3	US-09-199-637A-244
20	17.6	68.0	534	3	US-09-199-637A-246
21	17.6	68.0	601	3	US-09-949-016-74479
22	17.6	68.0	601	3	US-09-949-016-190379
23	17.6	68.0	42235	3	US-09-199-637A-1
24	17.6	68.0	69813	3	US-09-949-016-12455

c 25	17	68.0	69813	3	US-09-949-016-13905	Sequence 13905, A
c 26	17	68.0	69813	3	US-09-949-016-13906	Sequence 13906, A
c 27	17	68.0	69813	3	US-09-949-016-12861	Sequence 12861, A
c 28	17	68.0	79824	3	US-09-949-016-13919	Sequence 13919, A
c 29	17	68.0	148567	3	US-09-801-876B-3	Sequence 3, Appli
c 30	17	68.0	148567	3	US-10-254-863-3	Sequence 3, Appli
c 31	17	68.0	148567	3	US-10-667-442-3	Sequence 3, Appli
c 32	17	68.0	255679	3	US-09-949-016-17189	Sequence 17189, A
c 33	16.8	67.2	429	3	US-09-252-991A-14962	Sequence 14962, A
c 34	16.8	67.2	828	3	US-09-252-991A-15086	Sequence 15086, A
c 35	16.8	67.2	1347	3	US-09-252-991A-14827	Sequence 14827, A
c 36	16.8	67.2	146085	3	US-09-949-016-12872	Sequence 12872, A
c 37	16.8	67.2	146104	3	US-09-949-016-12339	Sequence 12339, A
c 38	16.6	66.4	601	3	US-09-949-016-35413	Sequence 35413, A
c 39	16.6	66.4	601	3	US-09-949-016-77995	Sequence 77995, A
c 40	16.6	66.4	601	3	US-09-949-016-196973	Sequence 196973, A
c 41	16.6	66.4	601	3	US-09-949-016-198113	Sequence 198113, A
c 42	16.6	66.4	1061	3	US-09-270-767-10689	Sequence 10689, A
c 43	16.6	66.4	1123	2	US-08-458-023B-3	Sequence 3, Appli
c 44	16.6	66.4	6676	3	US-09-573-080A-211	Sequence 211, App
c 45	16.6	66.4	9277	3	US-09-949-016-14017	Sequence 14017, A
c 46	16.6	66.4	13011	2	US-08-791-849A-14	Sequence 14, Appl
c 47	16.6	66.4	21168	3	US-09-949-016-12643	Sequence 12643, A
c 48	16.6	66.4	26334	3	US-09-949-016-17354	Sequence 17354, A
c 49	16.6	66.4	28843	3	US-09-949-016-17325	Sequence 17325, A
c 50	16.6	66.4	34531	3	US-09-949-016-14604	Sequence 14604, A
c 51	16.6	66.4	325034	3	US-09-949-016-14957	Sequence 14957, A
c 52	16.6	66.4	389504	3	US-09-949-016-11774	Sequence 11774, A
c 53	16.6	66.4	1230025	3	US-09-198-452A-1	Sequence 1, Appli
c 54	16.6	66.4	1230230	3	US-09-438-185A-1	Sequence 1, Appli
c 55	16.2	64.8	491	2	US-08-133-711-36	Sequence 36, Appli
c 56	16.2	64.8	601	3	US-09-949-002-2135	Sequence 2135, Ap
c 57	16.2	64.8	601	3	US-09-949-002-2136	Sequence 2136, Ap
c 58	16.2	64.8	601	3	US-09-949-002-8498	Sequence 8498, Ap
c 59	16.2	64.8	601	3	US-09-949-002-8499	Sequence 8499, Ap
c 60	16.2	64.8	718	3	US-09-154-083-31	Sequence 31, Appl
c 61	16.2	64.8	1008	3	US-09-101-423B-3	Sequence 3, Appli
c 62	16.2	64.8	1187	2	US-08-474-177-4	Sequence 4, Appli
c 63	16.2	64.8	1187	2	US-08-487-033-4	Sequence 4, Appli
c 64	16.2	64.8	1187	2	US-08-480-810-4	Sequence 4, Appli
c 65	16.2	64.8	1187	2	US-08-508-735-4	Sequence 4, Appli
c 66	16.2	64.8	1187	2	US-08-848-251-4	Sequence 4, Appli
c 67	16.2	64.8	1187	2	US-08-486-047-4	Sequence 4, Appli
c 68	16.2	64.8	1187	2	US-09-120-130-4	Sequence 4, Appli
c 69	16.2	64.8	1187	3	US-09-115-252-4	Sequence 4, Appli
c 70	16.2	64.8	1187	3	US-08-986-515-4	Sequence 4, Appli
c 71	16.2	64.8	1187	3	US-09-120-128-4	Sequence 4, Appli
c 72	16.2	64.8	1187	3	US-09-120-129-4	Sequence 4, Appli
c 73	16.2	64.8	1187	3	US-09-201-139-4	Sequence 4, Appli
c 74	16.2	64.8	1187	3	US-09-120-131-4	Sequence 4, Appli
c 75	16.2	64.8	1259	3	US-09-809-545A-40	Sequence 40, Appl
c 76	16.2	64.8	1259	3	US-10-623-629-4	Sequence 4, Appli
c 77	16.2	64.8	2311	3	US-10-104-047-1412	Sequence 1412, Ap
c 78	16.2	64.8	3205	3	US-09-061-768A-3	Sequence 3, Appli
c 79	16.2	64.8	3205	3	US-09-764-246-3	Sequence 3, Appli
c 80	16.2	64.8	3605	3	US-09-998-901-1	Sequence 1, Appli
c 81	16.2	64.8	5039	3	US-09-386-816C-1	Sequence 1, Appli
c 82	16.2	64.8	5039	3	US-10-320-176-1	Sequence 1, Appli
c 83	16.2	64.8	5342	3	US-09-807-757C-1	Sequence 1, Appli
c 84	16.2	64.8	5889	3	US-09-402-929-3	Sequence 3, Appli
c 85	16.2	64.8	7874	3	US-09-780-175-96	Sequence 96, Appl
c 86	16.2	64.8	9196	2	US-09-971-773-3	Sequence 3, Appli
c 87	16.2	64.8	12537	2	US-08-611-280-4	Sequence 4, Appli
c 88	16.2	64.8	12537	3	US-09-195-940-4	Sequence 4, Appli
c 89	16.2	64.8	12537	3	US-09-562-466-4	Sequence 4, Appli
c 90	16.2	64.8	14507	3	US-08-785-150-1	Sequence 1, Appli
c 91	16.2	64.8	14507	3	US-09-660-299-1	Sequence 1, Appli
c 92	16.2	64.8	14507	3	US-09-435-377-1	Sequence 1, Appli
c 93	16.2	64.8	14507	3	US-09-973-928-1	Sequence 1, Appli
c 94	16.2	64.8	17056	3	US-09-345-041-3	Sequence 3, Appli
c 95	16.2	64.8	17056	3	US-09-358-055B-3	Sequence 3, Appli
c 96	16.2	64.8	17056	3	US-09-893-238-3	Sequence 3, Appli
c 97	16.2	64.8	24593	3	US-09-949-016-13433	Sequence 13433, A

```

c 98 16.2 64.8 26700 2 US-08-472-217-1
c 99 16.2 64.8 26700 2 US-08-488-199-5
c 100 16.2 64.8 26700 3 US-08-760-534A-1
c 101 16.2 64.8 26700 3 US-09-336-757-1
c 102 16.2 64.8 30310 3 US-09-657-346A-96
c 103 16.2 64.8 90050 3 US-09-245-041-5
c 104 16.2 64.8 90050 3 US-09-358-055B-5
c 105 16.2 64.8 90050 3 US-09-893-238-5
c 106 16.2 64.8 90923 3 US-09-949-002-623
c 107 16.2 64.8 90925 3 US-09-949-002-789
c 108 16.2 64.8 134008 3 US-09-949-016-13841
c 109 16.2 64.8 264358 3 US-09-949-016-15725
c 110 16 64.0 319 3 US-09-513-998C-35607
c 111 16 64.0 441 3 US-09-513-999C-648
c 112 16 64.0 601 3 US-09-949-016-18039
c 113 16 64.0 601 3 US-09-949-016-19100
c 114 16 64.0 601 3 US-09-949-016-19101
c 115 16 64.0 601 3 US-09-949-016-59814
c 116 16 64.0 601 3 US-09-949-016-59815
c 117 16 64.0 601 3 US-09-949-016-59816
c 118 16 64.0 601 3 US-09-949-016-59817
c 119 16 64.0 601 3 US-09-949-016-68244
c 120 16 64.0 601 3 US-09-949-016-68245
c 121 16 64.0 601 3 US-09-949-016-68246
c 122 16 64.0 601 3 US-09-949-016-70266
c 123 16 64.0 601 3 US-09-949-016-145149
c 124 16 64.0 601 3 US-09-949-016-157744
c 125 16 64.0 601 3 US-09-949-016-157745
c 126 16 64.0 625 3 US-09-533-559-3038
c 127 16 64.0 944 3 US-09-270-767-15167
c 128 16 64.0 1163 3 US-09-160-588-1
c 129 16 64.0 1750 3 US-09-684-403-3
c 130 16 64.0 1833 2 US-08-365-470-2
c 131 16 64.0 2886 3 US-09-774-528-165
c 132 16 64.0 2886 3 US-10-120-988-165
c 133 16 64.0 3834 3 US-09-209-668-18
c 134 16 64.0 3834 3 US-09-949-016-71
c 135 16 64.0 3854 2 US-08-365-470-1
c 136 16 64.0 3858 2 US-08-344-155C-98
c 137 16 64.0 3858 3 US-09-009-490A-88
c 138 16 64.0 3858 3 US-09-982-262C-89
c 139 16 64.0 3862 3 US-09-949-016-4401
c 140 16 64.0 3863 3 US-08-482-073-1
c 141 16 64.0 3863 9 5217870-1
c 142 16 64.0 4286 3 US-08-227-800A-2
c 143 16 64.0 4286 3 US-08-921-954-2
c 144 16 64.0 6102 3 US-09-949-016-2007
c 145 16 64.0 10095 3 US-09-949-016-16141
c 146 16 64.0 14566 3 US-09-949-016-16765
c 147 16 64.0 15384 3 US-09-949-016-11813
c 148 16 64.0 15413 3 US-09-949-016-16143
c 149 16 64.0 18149 3 US-09-949-016-13204
c 150 16 64.0 23849 3 US-09-949-016-13813

```

ALIGNMENTS

```

RESULT 1
US-09-465-978A-51
; Sequence 51, Application US/09465978A
; Patent No. 6867348
; GENERAL INFORMATION:
; APPLICANT: Zhang, Nine and Anthony Purchio
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING FOR ANGIOGENESIS
; FILE REFERENCE: EXE-012.US
; CURRENT APPLICATION NUMBER: US/09/465,978A
; CURRENT FILING DATE: 1999-12-16
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 51
; LENGTH: 11176
; TYPE: DNA
US-09-465-978A-51
Query Match 71.2%; Score 17.8; DB 3; Length 11176;
Best Local Similarity 90.5%; Pred. No. 1,1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 5 AAGAGCGCAGACTGCTCTTCC 25
Db 10526 AAGACACAGACTGCTCTTCC 10546
RESULT 2
US-09-949-016-72599
; Sequence 72599, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72599
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-72599
Query Match 70.4%; Score 17.6; DB 3; Length 601;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAGAGCGCAGACTGCTCTTC 24
Db 573 AAAAAAGACATAGACTGCTTCC 596
RESULT 3
US-09-999-833A-528/c
; Sequence 528, Application US/09999833A
; Patent No. 6916648
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.

```

APPLICANT: Pan, James;  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2630PIC65  
CURRENT APPLICATION NUMBER: US/09/999,833A  
CURRENT FILING DATE: 2001-10-24  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639

```
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match          70.4%; Score 17.6; DB 3; Length 1285;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2  AAGAAGCGCAGCTGGTCTTCC 25
      | | | | | | | | | | | | | | | |
Db      816 AAGAAGAGCCGACGCTGGTCCACC 793

RESULT 4
US-10-020-445A-528/c
; Sequence 528, Application US/10020445A
; Patent No. 6962797
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2630PIC74
; CURRENT FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
```



/	PRIOR APPLICATION NUMBER:	60/084627
/	PRIOR FILING DATE:	1998-05-07
/	PRIOR APPLICATION NUMBER:	60/084643
/	PRIOR FILING DATE:	1998-05-07
/	PRIOR APPLICATION NUMBER:	60/085339
/	PRIOR FILING DATE:	1998-05-13
/	PRIOR APPLICATION NUMBER:	60/085338
/	PRIOR FILING DATE:	1998-05-13
/	PRIOR APPLICATION NUMBER:	60/085323
/	PRIOR FILING DATE:	1998-05-13
/	PRIOR APPLICATION NUMBER:	60/085582
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085700
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085689
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085579
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085580
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085573
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085704
/	PRIOR FILING DATE:	1998-05-15
/	PRIOR APPLICATION NUMBER:	60/085697

  

Query Match	70.4%;	Score 17
Best Local Similarity	83.3%;	Pred. No.
Matches	20; Conservative	* 0; Mismatches

  

Qy	2	AAAAGAGCCGACACTGGTCTTC	25
Db	816	AGAAAGGCCGACACTGGTCCACC	793

  

RESULT 5

US-09-999-833A-525/c

; Sequence 525, Application US/09999833A

; Patent No. 6916648

; GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Baker Kevin P.

; APPLICANT: Boenstein, David

; APPLICANT: Deenoyers, Luc

; APPLICANT: Eaton, Dan

; APPLICANT: Ferrara, Napoleon

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerber, Hanspeter

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, J Christopher

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Kijavlin, Ivar J.

; APPLICANT: Kuo, Sophia S.

; APPLICANT: Napier, Mary A.

; APPLICANT: Pan, James;

; APPLICANT: Paoni, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Shelton, David L.

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William I.

; TITLE OF INVENTION: Secrated and Transmembrane Acids Encoding the

; FILE REFERENCE: P2630P1C65

; CURRENT APPLICATION NUMBER: US/09/999,833

; CURRENT FILING DATE: 2001-10-24

; PRIOR APPLICATION NUMBER: 09/918585

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

```

; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1997-11-13

Query Match 70.4%; Score 17.6; DB 3; Length 2602;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAGAGCGCAGACTGGTCTCC 25
Db 785 AGAAGAGCCAGACTGGTCCACC 762

RESULT 6
US-10-020-445A-525/c
Sequence 525, Application US/10020445A
Patent No. 6962797
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Deenoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C74
CURRENT APPLICATION NUMBER: US/10/020,445A
CURRENT FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
```

; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/081952  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/081838  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082568  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082569  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082704  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082804  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082700  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082796  
; PRIOR FILING DATE: 1998-04-23  
; PRIOR APPLICATION NUMBER: 60/083336  
; PRIOR FILING DATE: 1998-04-27  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083392  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083496  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083499  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083545  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083554  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083558  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083559  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083500  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083742  
; PRIOR FILING DATE: 1998-04-30  
; PRIOR APPLICATION NUMBER: 60/084366  
; PRIOR FILING DATE: 1998-05-05  
; PRIOR APPLICATION NUMBER: 60/084414  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/084441  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/084637  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084639  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084640  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084598  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084627  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084643  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/085339  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/085338  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/085323  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/085592  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085700  
; PRIOR FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 60/085689  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085580  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085573  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085704  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085697

Query Match 70.4%; Score 17.6; DB 3; Length 2602;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGCAGACTGGTCTTCC 25  
Db 785 AGAAAGAGCCAGACTGGTCCACC 762

RESULT 7

US-09-667-373-3/c  
; Sequence 3, Application US/09667373  
; Patent No. 6524840  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Nehls, Michael C.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. 6524840el Human Endothelin Converting  
; FILE REFERENCE: LEX-0043-USA  
; CURRENT APPLICATION NUMBER: US/09/667,373  
; CURRENT FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US 60/156,102  
; PRIOR FILING DATE: 1999-09-24  
; PRIOR APPLICATION NUMBER: US 60/176,689  
; PRIOR FILING DATE: 2000-01-18  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2652  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-667-373-3

Query Match 70.4%; Score 17.6; DB 3; Length 2652;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAGAGCGCAGACTGGTCTTCC 25  
Db 1127 AGAAAGAGCCAGACTGGTCCACC 1104

RESULT 8

US-10-305-413-3/c  
; Sequence 3, Application US/10305413  
; Patent No. 6835564  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Nehls, Michael C.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. 6835564el Human Endothelin Converting  
; FILE REFERENCE: LEX-0043-USA  
; CURRENT APPLICATION NUMBER: US/10/305,413

```
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US 60/156,102
; PRIOR FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: US 60/176,689
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-305-413-3

Query Match          70.4%; Score 17.6; DB 3; Length 2652;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAAAAGCGCAGACTGGTCTTCC 25
Db      1127 AGAAAGAGCCAGACTGGTCTTCC 1104

RESULT 9
US-08-574-763-1/c
; Sequence 1, Application US/08574763
; Patent No. 5736376
; GENERAL INFORMATION:
; APPLICANT: Yanagisawa, Masashi
; TITLE OF INVENTION: RECOMBINANT ENDOTHELIN CONVERTING
; TITLE OF INVENTION: ENZYME-2 AND ITS USE IN ECE INHIBITOR SCREENING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574,763
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Corder, Timothy S.
; REGISTRATION NUMBER: 38,414
; REFERENCE/DOCKET NUMBER: UTXD:472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3291 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 116..2476
US-08-574-763-1

Query Match          70.4%; Score 17.6; DB 2; Length 3291;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAAAAGCGCAGACTGGTCTTCC 25
Db      1127 AGAAAGAGCCAGACTGGTCTTCC 1104

; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US 60/156,102
; PRIOR FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: US 60/176,689
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-305-413-3

Query Match          70.4%; Score 17.6; DB 3; Length 2652;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAAAAGCGCAGACTGGTCTTCC 25
Db      1127 AGAAAGAGCCAGACTGGTCTTCC 1104

RESULT 10
US-09-819-989-1/c
; Sequence 1, Application US/09819989
; Patent No. 6482629
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200
; CURRENT APPLICATION NUMBER: US/09/819,989
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3377
; TYPE: DNA
; ORGANISM: Human
US-09-819-989-1

Query Match          70.4%; Score 17.6; DB 3; Length 3377;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAAAAGCGCAGACTGGTCTTCC 25
Db      1024 AGAAAGAGCCAGACTGGTCTTCC 1001

RESULT 11
US-10-273-992-1/c
; Sequence 1, Application US/10273992
; Patent No. 6664093
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200DIV
; CURRENT APPLICATION NUMBER: US/10/273,992
; CURRENT FILING DATE: 2002-10-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3377
; TYPE: DNA
; ORGANISM: Human
US-10-273-992-1

Query Match          70.4%; Score 17.6; DB 3; Length 3377;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 AAAAAAGCGCAGACTGGTCTTCC 25
Db      1024 AGAAAGAGCCAGACTGGTCTTCC 1001

RESULT 12
US-10-681-222-1/c
; Sequence 1, Application US/10681222
; Patent No. 6825025
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200-DIV II
; CURRENT APPLICATION NUMBER: US/10/681,222
; CURRENT FILING DATE: 2003-10-09
```

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3377
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-681-222-1

Query Match          70.4%; Score 17.6; DB 3; Length 3377;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2  AAAAAGAGCGCAGACTGGTCTTCC 25
Db   1024 AGAAAGAGCCAGACTGGTCCACC 1001

RESULT 13
US-09-819-989-3/c
; Sequence 3, Application US/09819989
; Patent No. 6482629
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200
; CURRENT APPLICATION NUMBER: US/09/819,989
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 19650
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(19650)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-989-3

Query Match          70.4%; Score 17.6; DB 3; Length 19650;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2  AAAAAGAGCGCAGACTGGTCTTCC 25
Db   4443 AGAAAGAGCCAGACTGGTCCACC 4420

RESULT 14
US-10-273-992-3/c
; Sequence 3, Application US/10273992
; Patent No. 6664093
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200DIV
; CURRENT APPLICATION NUMBER: US/10/273,992
; CURRENT FILING DATE: 2002-10-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 19650
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(19650)
; OTHER INFORMATION: n = A,T,C or G
US-10-273-992-3
```

```
Query Match          70.4%; Score 17.6; DB 3; Length 19650;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2  AAAAAGAGCGCAGACTGGTCTTCC 25
Db   4443 AGAAAGAGCCAGACTGGTCCACC 4420

RESULT 15
US-10-681-222-3/c
; Sequence 3, Application US/10681222
; Patent No. 6825025
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001200-DIV II
; CURRENT APPLICATION NUMBER: US/10/681,222
; CURRENT FILING DATE: 2003-10-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 19650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(19650)
; OTHER INFORMATION: n = A,T,C or G
US-10-681-222-3

Query Match          70.4%; Score 17.6; DB 3; Length 19650;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2  AAAAAGAGCGCAGACTGGTCTTCC 25
Db   4443 AGAAAGAGCCAGACTGGTCCACC 4420

RESULT 16
US-09-949-016-13870/c
; Sequence 13870, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13870
; LENGTH: 155266
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13870

Query Match          70.4%; Score 17.6; DB 3; Length 155266;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  1  AAAAAGAGCGCAGACTGGTCTTC 24
```

Db 93329 AAAAAAGACATAGACTGGTCTCC 93306  
|||||||

RESULT 17  
US-09-949-016-170047  
; Sequence 170047, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 170047  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-170047

Query Match 68.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGGTCT 22  
|||||||  
Db 322 AAAAAAGAAAGCAGTCTGGTCT 343  
|||||||

RESULT 18  
US-09-949-016-16538/c  
; Sequence 16538, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16538  
; LENGTH: 13743  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16538

Query Match 68.8%; Score 17.2; DB 3; Length 13743;  
Best Local Similarity 86.4%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGGTCT 22  
|||||||  
Db 4472 AAAAAAGAAAGCAGTCTGGTCT 4451  
|||||||

RESULT 19

US-09-199-637A-244  
; Sequence 244, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 244  
; LENGTH: 501  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-244

Query Match 68.0%; Score 17; DB 3; Length 501;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||||  
Db 253 AAAAAAGCGCGCCAGATCTTCC 277  
|||||||

RESULT 20  
US-09-199-637A-246/c  
; Sequence 246, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 246  
; LENGTH: 534  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-246

Query Match 68.0%; Score 17; DB 3; Length 534;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||||  
Db 297 AAAAAAGCGCGCCAGATCTTCC 273  
|||||||



```
RESULT 21
US-09-949-016-74479/c
; Sequence 74479, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74479
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-74479

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
Db 231 AAAAAAGAGAAGAGCTGGTCTTAC 207

RESULT 22
US-09-949-016-190379/c
; Sequence 190379, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190379
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-190379

Query Match      68.0%; Score 17; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
Db 139 AAAAAAAGCGCAGACTGATCTGCC 115

RESULT 23
US-09-199-637A-1
; Sequence 1, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Rui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 42235
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(42235)
; OTHER INFORMATION: N is any nucleic acid.
US-09-199-637A-1

Query Match      68.0%; Score 17; DB 3; Length 42235;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
Db 18731 AAAAAAAGCGCCCGACAGATCTTCC 18755

RESULT 24
US-09-949-016-12455/c
; Sequence 12455, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12455
; LENGTH: 69813
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12455

Query Match      68.0%; Score 17; DB 3; Length 69813;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCTTCC 25
Db 4997 AAAAAAAGCTCAAACTTGCCTTCC 4973

RESULT 25
US-09-949-016-13905/c
; Sequence 13905, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13905
; LENGTH: 69813
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13905

Query Match          68.0%; Score 17; DB 3; Length 69813;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25
Db 4997 AAAAAAAGCTCAAACTTGCCTTCC 4973

RESULT 26
US-09-949-016-13906/c
; Sequence 13906, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13906
; LENGTH: 69813
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13906

Query Match          68.0%; Score 17; DB 3; Length 69813;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25
Db 4997 AAAAAAAGCTCAAACTTGCCTTCC 4973

RESULT 27
US-09-949-016-12861/c
; Sequence 12861, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001160
; CURRENT APPLICATION NUMBER: US/09/801,876B
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12861
; LENGTH: 69833
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12861

Query Match          68.0%; Score 17; DB 3; Length 69833;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25
Db 5017 AAAAAAAGCTCAAACTTGCCTTCC 4993

RESULT 28
US-09-949-016-13919/c
; Sequence 13919, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13919
; LENGTH: 79824
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13919

Query Match          68.0%; Score 17; DB 3; Length 79824;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25
Db 2173 AAAAAAGAGAAGACTGGTCTTAC 2149

RESULT 29
US-09-801-876B-3/c
; Sequence 3, Application US/09801876B
; Patent No. 6492155
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; FILE REFERENCE: CL001160
; CURRENT APPLICATION NUMBER: US/09/801,876B
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 148567
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(148567)
; OTHER INFORMATION: n = A,T,C or G
US-09-801-876B-3

Query Match      68.0%; Score 17; DB 3; Length 148567;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25
   ||||| ||||| ||||| |||||
Db 44191 AAAAAAGAGCTAAAGCTGTTCTTCC 44167

RESULT 30
US-10-254-869-3/c
; Sequence 3, Application US/10254869
; Patent No. 6653117
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001160DIV
; CURRENT APPLICATION NUMBER: US/10/254,869
; CURRENT FILING DATE: 2002-09-26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 148567
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(148567)
; OTHER INFORMATION: n = A,T,C or G
US-10-254-869-3

Query Match      68.0%; Score 17; DB 3; Length 148567;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25
   ||||| ||||| ||||| |||||
Db 44191 AAAAAAGAGCTAAAGCTGTTCTTCC 44167

RESULT 31
US-10-667-442-3/c
; Sequence 3, Application US/10667442
; Patent No. 6821765
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001160DIV II
; CURRENT APPLICATION NUMBER: US/10/667,442
; CURRENT FILING DATE: 2003-09-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 148567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

---

```
; LOCATION: (1)...(148567)
; OTHER INFORMATION: n = A,T,C or G
US-10-667-442-3

Query Match      68.0%; Score 17; DB 3; Length 148567;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25
   ||||| ||||| ||||| |||||
Db 44191 AAAAAAGAGCTAAAGCTGTTCTTCC 44167

RESULT 32
US-09-949-016-17189/c
; Sequence 17189, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17189
; LENGTH: 255679
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(255679)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17189

Query Match      68.0%; Score 17; DB 3; Length 255679;
Best Local Similarity 80.0%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGCTCTTCC 25
   ||||| ||||| ||||| |||||
Db 236732 AAAAAAGAGCGAGCTGATCTGCC 236708

RESULT 33
US-09-252-991A-14962
; Sequence 14962, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14962
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14962
```

```
Query Match          67.2%; Score 16.8; DB 3; Length 429;
Best Local Similarity 90.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAGAGCGCAGACTGCTT 23
DB 273 AAGGAGCGCAGACTGGTCAT 292

RESULT 34
US-09-252-991A-15086
; Sequence 15086, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15086
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15086

Query Match          67.2%; Score 16.8; DB 3; Length 828;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAGAGCGCAGACTGCTT 23
DB 344 AAGGAGCGCAGACTGGTCAT 363

RESULT 35
US-09-252-991A-14827
; Sequence 14827, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14827
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14827

Query Match          67.2%; Score 16.8; DB 3; Length 1347;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAGAGCGCAGACTGCTT 23
DB 562 AAGGAGCGCAGACTGGTCAT 581

RESULT 36
US-09-949-016-12872/c
```

```
; Sequence 12872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12872
; LENGTH: 146095
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12872
```

```
Query Match          67.2%; Score 16.8; DB 3; Length 146095;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAGAGCGCAGACTGGT 20
DB 66917 AGAAACAGCGCAGACTGGT 66898
```

```
RESULT 37
US-09-949-016-13239/c
; Sequence 13239, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13239
; LENGTH: 146104
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13239
```

```
Query Match          67.2%; Score 16.8; DB 3; Length 146104;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAGAGCGCAGACTGGT 20
DB 66917 AGAAACAGCGCAGACTGGT 66898
```

```
RESULT 38
US-09-949-016-35413/c
; Sequence 35413, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35413
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-35413

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTT 23
   ||||| ||||| ||||| ||||| |||||
Db 236 AAAAAAAGCGCAGACTTGT 214

RESULT 39
US-09-949-016-77995/c
; Sequence 77995, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77995
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-77995

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTT 23
   ||||| ||||| ||||| ||||| |||||
Db 236 AAAAAAAGCGCAGACTTGT 214

RESULT 40
US-09-949-016-196973/c
; Sequence 196973, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 196973
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-196973

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAAGCGCAGACTGGTCTTCC 25
   ||||| ||||| ||||| ||||| |||||
Db 499 AAACGCTGCAGACTGGTCTTCC 477

RESULT 41
US-09-949-016-198113/c
; Sequence 198113, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 198113
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-198113

Query Match          66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAAGCGCAGACTGGTCTTCC 25
   ||||| ||||| ||||| ||||| |||||
Db 188 AAAATAGCCAGACTGTTTTC 166

RESULT 42
US-09-270-767-10689/c
; Sequence 10689, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10689
; LENGTH: 1061
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-10689
```



QY 1 AAAAAAGCGCAGACTGGTCTT 23  
 ||||| ||||| ||||| |||||  
 Db 3212 AAAAAAGCGCAGACTGGTTT 3190

RESULT 46  
 US-08-791-849A-14/c  
 ; Sequence 14, Application US/08791849A  
 ; Patent No. 5914449  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makoto MURASE et al.  
 ; TITLE OF INVENTION: Method for Increasing Storage  
 ; TITLE OF INVENTION: Lipid Content in Plant Seed  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wenderoth, Lind & Ponack  
 ; STREET: 805 Fifteenth Street, N.W., #700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: Wordperfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/791,849A  
 ; FILING DATE: January 30, 1997  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warren M. Cheek, Jr.  
 ; REGISTRATION NUMBER: 33,367  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-8850  
 ; TELEFAX:  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 13011 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: rat (Rattus norvegicus)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: join(3212..3218, 3766..3948, 5917..6008,  
 ; LOCATION: 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8077  
 ; LOCATION: 9298..9479, 10163..10269)  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: join(3194..3218, 3766..3948, 5917..6008,  
 ; LOCATION: 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8077  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: join(3219..3765, 3949..5916, 6009..6151,  
 ; LOCATION: 6284..6417, 6605..6920, 7191..7301, 7453..7664, 7818..7910, 8078..9297  
 ; LOCATION: 9480..10162)  
 US-08-791-849A-14

Query Match 66.4%; Score 16.6; DB 2; Length 13011;  
 Best Local Similarity 82.8%; Pred. No. 4.2e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 AAAAAAGCGCAGACTGGTCTTCC 25  
 ||||| ||||| ||||| |||||

Db 9788 ATAAGACACTGACTGTTCTTCC 9766  
 RESULT 47  
 US-09-949-016-12643/c  
 ; Sequence 12643, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12643  
 ; LENGTH: 21168  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-12643

Query Match 66.4%; Score 16.6; DB 3; Length 21168;  
 Best Local Similarity 82.6%; Pred. No. 4.5e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTT 23  
 ||||| ||||| ||||| |||||  
 Db 3214 AAAAAAGCGCAGACTGGTTT 3192

RESULT 48  
 US-09-949-016-17354/c  
 ; Sequence 17354, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17354  
 ; LENGTH: 26334  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-17354

Query Match 66.4%; Score 16.6; DB 3; Length 26334;  
 Best Local Similarity 82.6%; Pred. No. 4.7e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAAGCGCAGACTGGTCTTCC 25  
 ||||| ||||| ||||| |||||  
 Db 10129 AAAAAAGCGCAGACTGGTTTCC 10107

RESULT 49  
 US-09-949-016-17325



Search completed: February 3, 2006, 16:32:55  
Job time : 86.6667 secs

```
; Sequence 17325, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17325
; LENGTH: 28843
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(28843)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17325
```

```
Query Match 66.4%; Score 16.6; DB 3; Length 28843;
Best Local Similarity 82.6%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 AAAAGCGGCGAGCTGGTCTTCC 25
Db 9644 AAACGCTGGCAGACTGGTCTTCC 9666
```

```
RESULT 50
US-09-949-016-14604/C
; Sequence 14604, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14604
; LENGTH: 34531
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(34531)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14604
```

```
Query Match 66.4%; Score 16.6; DB 3; Length 34531;
Best Local Similarity 82.6%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAGAGCGCAGCTGGTCTT 23
Db 17328 AAAAAGAGAGAGCTGGTCTT 17306
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 14:37:28 ; Search time 365.556 Seconds  
(without alignments)  
565.535 Million cell updates/sec

Title: US-10-719-900-35

Perfect score: 25

Sequence: 1 aaaaagagcgactgtcttc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA\_Main:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	8	US-10-719-900-35
2	23.4	93.6	25	8	US-10-719-900-36
3	22.4	89.6	25	8	US-10-719-900-342
4	21.8	87.2	36599	7	US-10-322-281-679
5	20.8	83.2	25	8	US-10-719-900-341
6	19.4	77.6	30752	5	US-10-087-192-745
7	19.4	77.6	90100	3	US-09-997-722-295
8	19.2	76.8	626	4	US-09-925-065A-496554
9	19.2	76.8	626	4	US-09-925-065A-496555
10	19.2	76.8	1276	5	US-10-243-820-6
11	19.2	76.8	1276	5	US-10-243-817-6
12	19.2	76.8	1392	7	US-10-282-122A-24223
13	19.2	76.8	29956	3	US-09-997-722-229
14	19.2	76.8	123192	5	US-10-175-523-71
15	19.2	76.8	123192	10	US-11-099-266-71
16	18.8	75.2	43445	8	US-10-712-795-334
17	18.8	75.2	43445	8	US-10-920-612-334
18	18.8	75.2	54775	7	US-10-741-601-5659
19	18.8	75.2	54775	8	US-10-741-600-17684
20	18.6	74.4	502	7	US-10-767-701-24439
21	18.6	74.4	667	5	US-10-027-632-234119
22	18.6	74.4	667	6	US-10-027-632-234119
23	18.6	74.4	769	5	US-10-027-632-162620

c	24	18.6	74.4	769	6	US-10-027-632-162620	Sequence 162620,
	25	18.6	74.4	1305	5	US-10-027-632-204028	Sequence 204028,
	26	18.6	74.4	1305	6	US-10-027-632-204028	Sequence 204028,
c	27	18.6	74.4	7656	9	US-10-621-911A-7	Sequence 7, Appli
c	28	18.6	74.4	7656	9	US-10-646-390A-7	Sequence 7, Appli
	29	18.6	74.4	39405	5	US-10-087-192-1285	Sequence 1285, Ap
	30	18.6	74.4	39790	5	US-10-087-192-1423	Sequence 1423, Ap
c	31	18.6	74.4	47115	7	US-10-052-482-133	Sequence 133, App
c	32	18.6	74.4	49281	5	US-10-087-192-301	Sequence 301, App
c	33	18.6	74.4	64482	7	US-10-322-696-151	Sequence 151, App
c	34	18.6	74.4	89213	7	US-10-322-281-843	Sequence 843, App
	35	18.4	73.6	47115	7	US-10-052-482-133	Sequence 133, App
	36	18.4	72.8	600	4	US-09-925-065A-835917	Sequence 835917,
	37	17.8	71.2	331	3	US-09-728-445-172	Sequence 172, App
	38	17.8	71.2	331	3	US-10-964-549-172	Sequence 172, App
	39	17.8	71.2	4333	6	US-10-359-050-17	Sequence 2, Appli
	40	17.8	71.2	5409	9	US-10-685-837-2	Sequence 18, Appli
	41	17.8	71.2	10491	6	US-10-359-050-18	Sequence 43, Appli
	42	17.8	71.2	11176	3	US-09-738-968-43	Sequence 20, Appli
	43	17.8	71.2	11784	6	US-10-359-050-20	Sequence 106, App
	44	17.8	71.2	11784	6	US-10-014-099F-106	Sequence 12, Appli
	45	17.8	71.2	12538	8	US-10-359-050-12	Sequence 1, Appli
	46	17.8	71.2	12607	8	US-10-789-465-1	Sequence 1, Appli
	47	17.8	71.2	12607	8	US-10-867-628-1	Sequence 13, Appli
	48	17.8	71.2	12645	6	US-10-359-050-13	Sequence 1, Appli
	49	17.8	71.2	13139	9	US-10-685-837-1	Sequence 4, Appli
	50	17.8	71.2	14947	9	US-10-685-837-4	Sequence 8, Appli
	51	17.8	71.2	15174	9	US-10-685-837-8	Sequence 6, Appli
	52	17.8	71.2	15199	9	US-10-685-837-6	Sequence 171, Ap
c	53	17.8	71.2	21635	5	US-10-087-192-1171	Sequence 97, Appli
	54	17.8	71.2	22114	7	US-10-322-281-97	Sequence 97, Appli
	55	17.8	71.2	25543	5	US-10-087-192-805	Sequence 805, App
	56	17.8	71.2	26666	8	US-10-087-192-1777	Sequence 1777, Ap
	57	17.8	71.2	28446	5	US-10-331-053-13	Sequence 13, Appli
c	58	17.8	71.2	30654	5	US-10-087-192-2035	Sequence 2035, Ap
	59	17.8	71.2	32591	6	US-10-085-117-55	Sequence 55, Appli
	60	17.8	71.2	35183	7	US-10-322-281-445	Sequence 445, App
	61	17.8	71.2	35359	6	US-10-087-192-421	Sequence 421, App
c	62	17.8	71.2	38646	5	US-10-085-117-193	Sequence 193, App
c	63	17.8	71.2	43591	5	US-10-087-192-1603	Sequence 1603, Ap
	64	17.8	71.2	44372	5	US-10-087-192-349	Sequence 349, App
	65	17.8	71.2	47299	7	US-10-322-281-243	Sequence 243, App
	66	17.8	71.2	49088	5	US-10-087-192-13	Sequence 13, Appli
c	67	17.8	71.2	49826	5	US-10-087-192-739	Sequence 739, App
c	68	17.8	71.2	54303	8	US-10-417-375-79	Sequence 79, Appli
	69	17.8	71.2	55544	5	US-10-087-192-811	Sequence 811, App
c	70	17.8	71.2	56773	8	US-10-331-053-47	Sequence 47, Appli
	71	17.8	71.2	58665	5	US-10-087-192-1975	Sequence 1975, Ap
	72	17.8	71.2	65559	5	US-10-087-192-841	Sequence 841, App
c	73	17.8	71.2	65952	7	US-10-087-192-841	Sequence 841, App
c	74	17.8	71.2	66126	7	US-10-322-696-163	Sequence 163, App
c	75	17.8	71.2	69515	5	US-10-322-281-423	Sequence 423, App
c	76	17.8	71.2	70215	5	US-10-087-192-1219	Sequence 1219, Ap
c	77	17.8	71.2	72815	5	US-10-087-192-217	Sequence 217, App
c	78	17.8	71.2	72815	7	US-10-322-281-265	Sequence 265, App
c	79	17.8	71.2	79356	7	US-10-322-281-167	Sequence 167, App
c	80	17.8	71.2	80423	7	US-10-367-094-41	Sequence 41, Appli
	81	17.8	71.2	84410	7	US-10-322-281-747	Sequence 747, App
	82	17.8	71.2	85435	6	US-10-085-117-127	Sequence 127, App
	83	17.8	71.2	88493	7	US-10-322-281-703	Sequence 703, App
c	84	17.8	71.2	90743	6	US-10-085-117-271	Sequence 271, App
	85	17.8	71.2	96389	3	US-10-052-482-181	Sequence 181, App
	86	17.8	71.2	96599	7	US-09-997-722-199	Sequence 199, App
	87	17.8	71.2	102145	8	US-10-331-053-53	Sequence 53, Appli
	88	17.8	71.2	111252	5	US-10-087-192-895	Sequence 895, App
c	89	17.8	71.2	117329	9	US-10-987-384-1	Sequence 1, Appli
	90	17.8	71.2	130427	5	US-10-175-523-87	Sequence 87, Appli
	91	17.8	71.2	130427	10	US-11-099-266-87	Sequence 87, Appli
c	92	17.8	71.2	155579	6	US-10-085-117-283	Sequence 283, App
	93	17.8	71.2	168976	6	US-10-085-117-139	Sequence 139, App
	94	17.8	71.2	171936	6	US-10-085-071-24	Sequence 24, Appli
	95	17.8	71.2	171936	6	US-10-025-966A-24	Sequence 24, Appli
	96	17.8	71.2	215980	3	US-09-972-546-16	Sequence 16, Appli



```
RESULT 4
US-10-322-281-679
; Sequence 679, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 679
; LENGTH: 36599
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(36599)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-679

Query Match      87.2%; Score 21.8; DB 7; Length 36599;
Best Local Similarity 92.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
Db 35597 AAAAAAGCGCAGACTGGTCTTCC 35621

RESULT 5
US-10-719-900-341
; Sequence 341, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
; FILE REFERENCE: 3528.1
; CURRENT APPLICATION NUMBER: US/10/719,900
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,808
; PRIOR FILING DATE: 2002.11.20
; NUMBER OF SEQ ID NOS: 982914
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 341
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-719-900-341

Query Match      83.2%; Score 20.8; DB 8; Length 25;
Best Local Similarity 91.7%; Pred. No. 4.8;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAGCGCAGACTGGTCTTCC 25
Db 1 AAAAAAGCGCTCAGTGGTCTTCC 24

RESULT 6
US-10-087-192-745
; Sequence 745, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 745
; LENGTH: 30752
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(30752)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-745

Query Match      77.6%; Score 19.4; DB 5; Length 30752;
Best Local Similarity 95.2%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 3215 AAGAGCGCAGACTGGTCTTCC 3235

RESULT 7
US-09-997-722-295
; Sequence 295, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 295
; LENGTH: 90100
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1696)..(5225)
; OTHER INFORMATION: "n" at positions 1696 through 5225 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10303)..(10322)
; OTHER INFORMATION: "n" at positions 10303 through 10322 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (32410)..(32429)
; OTHER INFORMATION: "n" at positions 32410 through 32429 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34679)..(35400)
; OTHER INFORMATION: "n" at positions 34679 through 35400 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (38214)..(43788)
; OTHER INFORMATION: "n" at positions 38214 through 43788 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45408)..(46631)
; OTHER INFORMATION: "n" at positions 45408 through 46631 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47453)..(47810)
```

OTHER INFORMATION: "n" at positions 47453 through 47810 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (49455)..(49572)  
OTHER INFORMATION: "n" at positions 49495 through 49572 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (52976)..(53745)  
OTHER INFORMATION: "n" at positions 52976 through 53745 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (59340)..(60268)  
OTHER INFORMATION: "n" at positions 59340 through 60268 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (76210)..(76409)  
OTHER INFORMATION: "n" at positions 76210 through 76409 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (83590)..(83609)  
OTHER INFORMATION: "n" at positions 83590 through 83609 can be any base.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (88024)..(88383)  
OTHER INFORMATION: "n" at positions 88024 through 88383 can be any base.  
US-09-997-722-295

Query Match 77.8%; Score 19.4; DB 3; Length 90100;  
Best Local Similarity 95.2%; Pred. No. 47;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 AAGAGCGCAGACTGGCTCTCC 25  
Db 50180 AAGAGCGCAGACTGGCTCTCC 50200

RESULT 8  
US-09-925-065A-496554  
; Sequence 496554, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 496554  
; LENGTH: 626  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-496554

Query Match 76.8%; Score 19.2; DB 4; Length 626;  
Best Local Similarity 87.5%; Pred. No. 38;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGAGCGCAGACTGGCTCTCC 24  
Db 284 AAAAAAGAGTCAGAAATGGCTCTCC 307

RESULT 9  
US-09-925-065A-496555  
; Sequence 496555, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 496555  
; LENGTH: 626  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-496555

Query Match 76.8%; Score 19.2; DB 4; Length 626;  
Best Local Similarity 87.5%; Pred. No. 38;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAGAGCGCAGACTGGCTCTCC 24  
Db 284 AAAAAAGAGTCAGAAATGGCTCTCC 307

RESULT 10  
US-10-243-820-6/c  
; Sequence 6, Application US/10243820  
; Publication No. US20030101471A1  
; GENERAL INFORMATION:  
; APPLICANT: BALTIMORE, DAVID  
; APPLICANT: HONG, ELIZABETH J.  
; APPLICANT: LOIS-CABALLE, CARLOS  
; APPLICANT: PEASE, SHIRLEY  
; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSGENIC ANIMALS  
; FILE REFERENCE: CALTE.009A  
; CURRENT APPLICATION NUMBER: US/10/243,820  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 60/322,031  
; PRIOR FILING DATE: 2001-09-13  
; PRIOR APPLICATION NUMBER: 60/347,782  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1276  
; TYPE: DNA  
; ORGANISM: Murine  
US-10-243-820-6

Query Match 76.8%; Score 19.2; DB 5; Length 1276;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAGAGCGCAGACTGGCTCTCC 25  
Db 66 AAAAAAGAGCACTGACTGCTCTCC 43

RESULT 11  
US-10-243-817-6/c  
; Sequence 6, Application US/10243817  
; Publication No. US20030101472A1  
; GENERAL INFORMATION:  
; APPLICANT: BALTIMORE, DAVID  
; APPLICANT: HONG, ELIZABETH J.  
; APPLICANT: LOIS-CABALLE, CARLOS  
; APPLICANT: PEASE, SHIRLEY  
; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSGENIC ANIMALS  
; FILE REFERENCE: CALTE.006A  
; CURRENT APPLICATION NUMBER: US/10/243,817  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 60/322,031  
; PRIOR FILING DATE: 2001-09-13  
; PRIOR APPLICATION NUMBER: 60/347,782  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1276  
; TYPE: DNA  
; ORGANISM: Murine  
US-10-243-817-6

Query Match 76.8%; Score 19.2; DB 5; Length 1276;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 66 AAAAAAGCAGCTGCTCTTCC 43

RESULT 12  
US-10-282-122A-24223/c  
; Sequence 24223, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24223  
; LENGTH: 1392  
; TYPE: DNA  
; ORGANISM: Listeria monocytogenes  
US-10-282-122A-24223

Query Match 76.8%; Score 19.2; DB 7; Length 1392;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 24  
|||||  
Db 132 AAAAAAGCGCGACTGATTTTC 109

RESULT 13  
US-09-997-722-229/c  
; Sequence 229, Application US/09997722  
; Publication No. US20040072154A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71171/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/09/997,722  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 301  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 229  
; LENGTH: 29956  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1335)..(1858)  
; OTHER INFORMATION: "n" at positions 1335 through 1858 can be any base.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (7091)..(7110)  
; OTHER INFORMATION: "n" at positions 7091 through 7110 can be any base.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (20085)..(20586)  
; OTHER INFORMATION: "n" at positions 20085 through 20586 can be any base.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (24372)..(24840)  
; OTHER INFORMATION: "n" at positions 24372 through 24840 can be any base.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29121)..(29589)  
; OTHER INFORMATION: "n" at positions 29121 through 29589 can be any base.  
US-09-997-722-229

Query Match 76.8%; Score 19.2; DB 3; Length 29956;  
Best Local Similarity 87.5%; Pred. No. 54;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 8863 AAAAAAGCAGCTGCTCTTCC 8840

RESULT 14



```
US-10-175-523-71
; Sequence 71, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/13795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 123192
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-71

Query Match          76.8%; Score 19.2; DB 5; Length 123192;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGCTCTTC 24
Db 73891 AAAAAAGAGCGACTGCTCTTC 73914

RESULT 15
US-11-099-266-71
; Sequence 71, Application US/11099266
; Publication No. US20050181433A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 03235/100J795-US4
; CURRENT APPLICATION NUMBER: US/11/099,266
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US 10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
```

```
US-10-920-612-334
; Sequence 334, Application US/10920612
; Publication No. US20050009088A1
; GENERAL INFORMATION:
; APPLICANT: Crooke et al.
; TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
; FILE REFERENCE: 30566/39634A
; CURRENT APPLICATION NUMBER: US/10/920,612
; CURRENT FILING DATE: 2004-08-17
; PRIOR APPLICATION NUMBER: PCT/US03/15493
; PRIOR FILING DATE: 2003-11-15
; PRIOR APPLICATION NUMBER: US 10/712,795
; PRIOR FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: US 60/426,234
; PRIOR FILING DATE: 2002-11-13
; NUMBER OF SEQ ID NOS: 892
; SEQ ID NO 334
; LENGTH: 43445
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-920-612-334

Query Match          75.2%; Score 18.8; DB 8; Length 43445;
Best Local Similarity 90.9%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGAGACTGCTCT 22
Db 25768 AAAACAGAGCGTAGACTGCTCT 25789

RESULT 17
US-10-920-612-334
; Sequence 334, Application US/10920612
; Publication No. US20050009088A1
; GENERAL INFORMATION:
; APPLICANT: Crooke et al.
; TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
; FILE REFERENCE: 30566/39634A
; CURRENT APPLICATION NUMBER: US/10/920,612
; CURRENT FILING DATE: 2004-08-17
; PRIOR APPLICATION NUMBER: PCT/US03/15493
; PRIOR FILING DATE: 2003-11-15
; PRIOR APPLICATION NUMBER: US 10/712,795
; PRIOR FILING DATE: 2003-11-13
; PRIOR APPLICATION NUMBER: US 60/426,234
; PRIOR FILING DATE: 2002-11-13
; NUMBER OF SEQ ID NOS: 892
; SEQ ID NO 334
; LENGTH: 43445
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-920-612-334

Query Match          75.2%; Score 18.8; DB 8; Length 43445;
Best Local Similarity 90.9%; Pred. No. 87;
```

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCT 22  
Db 25768 AAAAAAGCGCAGACTGGTCT 25789

## RESULT 18

US-10-741-601-5659  
; Sequence 5659, Application US/10741601  
; Publication No. US20040166519A1

; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001500  
; CURRENT APPLICATION NUMBER: US/10/741,601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 5659  
; LENGTH: 54775

; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-741-601-5659

Query Match 75.2%; Score 18.8; DB 7; Length 54775;  
Best Local Similarity 90.9%; Pred. No. 89;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCT 22  
Db 31373 AAAAAAGCGCAGACTGGTCT 31394

## RESULT 19

US-10-741-600-17684  
; Sequence 17684, Application US/10741600  
; Publication No. US20050026169A1

; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001499  
; CURRENT APPLICATION NUMBER: US/10/741,600  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 73997  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 17684  
; LENGTH: 54775

; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-741-600-17684

Query Match 75.2%; Score 18.8; DB 8; Length 54775;  
Best Local Similarity 90.9%; Pred. No. 89;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCT 22  
Db 31373 AAAAAAGCGCAGACTGGTCT 31394

## RESULT 20

US-10-767-701-24439/c  
; Sequence 24439, Application US/10767701  
; Publication No. US2004017684A1

; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 24439  
; LENGTH: 502

; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 30935687  
US-10-767-701-24439

Query Match 74.4%; Score 18.6; DB 7; Length 502;  
Best Local Similarity 84.0%; Pred. No. 74;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25  
Db 265 AAAAAAGCGCAGACTGGTCTTCC 241

## RESULT 21

US-10-027-632-234119  
; Sequence 234119, Application US/10027632  
; Publication No. US20020198371A1

; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 234119  
; LENGTH: 667  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-234119

Query Match 74.4%; Score 18.6; DB 5; Length 667;  
Best Local Similarity 84.0%; Pred. No. 76;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAGCGCAGACTGGTCTTCC 25  
Db 600 AAAAAAGCGCAGACTGGTCTTCC 624

## RESULT 22

US-10-027-632-234119  
; Sequence 234119, Application US/10027632  
; Publication No. US20030204075A9

; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632

RESULT 23  
US-10-027-632-162620/c  
; Sequence 162620, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 234119  
; LENGTH: 667  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-234119

Query Match 74.4%; Score 18.6; DB 6; Length 667;  
Best Local Similarity 84.0%; Pred. No. 76;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTTCC 25  
|||||||  
DB 600 AAAAAAGCGCAGACTGGCTTCC 624

RESULT 23  
US-10-027-632-162620/c  
; Sequence 162620, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 162620  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-162620

Query Match 74.4%; Score 18.6; DB 5; Length 769;  
Best Local Similarity 84.0%; Pred. No. 77;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTTCC 25  
|||||||  
DB 171 AAAAAAGCGCAGACTGGCTTCC 147

RESULT 24  
US-10-027-632-162620/c  
; Sequence 162620, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 162620  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-162620

Query Match 74.4%; Score 18.6; DB 6; Length 769;  
Best Local Similarity 84.0%; Pred. No. 77;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGCTTCC 25  
|||||||  
DB 171 AAAAAAGCGCAGACTGGCTTCC 147

RESULT 25  
US-10-027-632-204028  
; Sequence 204028, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 204028  
; LENGTH: 1305  
; TYPE: DNA

```

; ORGANISM: Human
US-10-027-632-204028

Query Match          74.4%; Score 18.6; DB 5; Length 1305;
Best Local Similarity 84.0%; Pred. No. 80;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| ||||| |||||
Db 518 AAAAAAGGTCAGACTTGGCTTCC 542

RESULT 26
US-10-027-632-204028
; Sequence 204028, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 204028
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-204028

Query Match          74.4%; Score 18.6; DB 6; Length 1305;
Best Local Similarity 84.0%; Pred. No. 80;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| ||||| |||||
Db 518 AAAAAAGGTCAGACTTGGCTTCC 542

RESULT 27
US-10-621-911A-7/c
; Sequence 7, Application US/10621911A
; Publication No. US20050054823A1
; GENERAL INFORMATION:
; APPLICANT: SAITOU, Mitinori
; TITLE OF INVENTION: Genes
; FILE REFERENCE: 674558-2002
; CURRENT APPLICATION NUMBER: US/10/621,911A
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: PCT/GB02/00215
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: GB 0101300.2
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 7
; LENGTH: 7656
; TYPE: DNA
; ORGANISM: Rattus sp
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7471)..(7471)
; OTHER INFORMATION: "n" is an unknown nucleotide
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7554)..(7554)
; OTHER INFORMATION: "n" is an unknown nucleotide
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7608)..(7608)
; OTHER INFORMATION: "n" is an unknown nucleotide
US-10-646-390A-7

Query Match          74.4%; Score 18.6; DB 9; Length 7656;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| ||||| |||||
Db 2113 AAAAAAGAGTGTGACTGTCTTCC 2089

RESULT 28
US-10-646-390A-7/c
; Sequence 7, Application US/10646390A
; Publication No. US20050054824A1
; GENERAL INFORMATION:
; APPLICANT: SAITOU, Mitinori
; TITLE OF INVENTION: Genes
; FILE REFERENCE: 674558-2002.1
; CURRENT APPLICATION NUMBER: US/10/646,390A
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: 10/621,911
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: PCT/GB02/00215
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: GB 0101300.2
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 7
; LENGTH: 7656
; TYPE: DNA
; ORGANISM: Rattus sp
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7471)..(7471)
; OTHER INFORMATION: "n" is an unknown nucleotide
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7554)..(7554)
; OTHER INFORMATION: "n" is an unknown nucleotide
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7608)..(7608)
; OTHER INFORMATION: "n" is an unknown nucleotide
US-10-646-390A-7

Query Match          74.4%; Score 18.6; DB 9; Length 7656;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| ||||| |||||
Db 2113 AAAAAAGAGTGTGACTGTCTTCC 2089

RESULT 29
US-10-087-192-1285
```

```
; Sequence 1285, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1285
; LENGTH: 39405
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(39405)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1285
```

```
Query Match 74.4%; Score 18.6; DB 5; Length 39405;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAGAGCGAGACTGCTCTTCC 25
Db 9488 AGAAGAGACATGACTGCTCTTCC 9512
```

```
RESULT 30
US-10-087-192-1423
; Sequence 1423, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1423
; LENGTH: 39790
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(39790)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1423
```

```
Query Match 74.4%; Score 18.6; DB 5; Length 39790;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAGAGCGAGACTGCTCTTCC 25
Db 15595 AATAAGAGACATGACTGCTCTTCC 15619
```

```
RESULT 31
```

```
US-10-052-482-133/c
; Sequence 133, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 133
; LENGTH: 47115
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1171)..(1270)
; OTHER INFORMATION: "n" at positions 1171 to 1270 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28066)..(28402)
; OTHER INFORMATION: "n" at positions 28066 to 28402 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34855)..(34891)
; OTHER INFORMATION: "n" at positions 34855 to 34891 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (37116)..(37389)
; OTHER INFORMATION: "n" at positions 37116 to 37389 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (43130)..(44280)
; OTHER INFORMATION: "n" at positions 43130 to 44280 can be any base
US-10-052-482-133
```

```
Query Match 74.4%; Score 18.6; DB 7; Length 47115;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAGAGCGAGACTGCTCTTCC 25
Db 5679 AAAAAAGAGCATGACTGCTCTTCC 5655
```

```
RESULT 32
US-10-087-192-301/c
; Sequence 301, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 49281
; TYPE: DNA
; ORGANISM: Mus musculus
```

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(49281)  
OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-301

Query Match 74.4%; Score 18.6; DB 5; Length 49281;  
Best Local Similarity 84.0%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 11432 AAAAAAGCGCAGACTGGTCTTCC 11408

RESULT 33  
US-10-322-696-151/c  
Sequence 151, Application US/10322696  
Publication No. US20040166490A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Malandro, Marc  
TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER  
FILE REFERENCE: 529452001200  
CURRENT APPLICATION NUMBER: US/10/322,696  
CURRENT FILING DATE: 2003-10-17  
NUMBER OF SEQ ID NOS: 186  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 151  
LENGTH: 64482  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(64482)  
OTHER INFORMATION: n = A,T,C or G  
US-10-322-696-151

Query Match 74.4%; Score 18.6; DB 7; Length 64482;  
Best Local Similarity 84.0%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 53861 AATAAGAGCAGTGCCTTCC 53837

RESULT 34  
US-10-322-281-843/c  
Sequence 843, Application US/10322281  
Publication No. US20040126762A1  
GENERAL INFORMATION:  
APPLICANT: David W. Morris  
APPLICANT: Marc S. Malandro  
TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
FILE REFERENCE: 529452001000  
CURRENT APPLICATION NUMBER: US/10/322,281  
CURRENT FILING DATE: 2002-12-17  
NUMBER OF SEQ ID NOS: 866  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 843  
LENGTH: 89213  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(89213)  
OTHER INFORMATION: n = A,T,C or G  
US-10-322-281-843

Query Match 74.4%; Score 18.6; DB 7; Length 89213;  
Best Local Similarity 84.0%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTCC 25  
|||||  
Db 81877 AAAAAAGAGTGTGAGCTGCTTCC 81853

RESULT 35  
US-10-052-482-133  
Sequence 133, Application US/10052482  
Publication No. US20040072264A1  
GENERAL INFORMATION:  
APPLICANT: Engelhard, Eric  
APPLICANT: Morris, David  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
FILE REFERENCE: A-71087/RMS/DCF  
CURRENT APPLICATION NUMBER: US/10/052,482  
CURRENT FILING DATE: 2002-08-15  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 241  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 133  
LENGTH: 47115  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1171)..(1270)  
OTHER INFORMATION: "n" at positions 1171 to 1270 can be any base  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28066)..(28402)  
OTHER INFORMATION: "n" at positions 28066 to 28402 can be any base  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (34855)..(34891)  
OTHER INFORMATION: "n" at positions 34855 to 34891 can be any base  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (37116)..(37389)  
OTHER INFORMATION: "n" at positions 37116 to 37389 can be any base  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (43130)..(44280)  
OTHER INFORMATION: "n" at positions 43130 to 44280 can be any base  
US-10-052-482-133

Query Match 73.6%; Score 18.4; DB 7; Length 47115;  
Best Local Similarity 95.0%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 24  
|||||  
Db 34115 AAGAGCGCAGACTGGTCTTCC 34134

RESULT 36  
US-09-925-065A-835917  
Sequence 835917, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
Nucleotide Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 835917  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-835917

Query Match 72.8%; Score 18.2; DB 4; Length 600;  
Best Local Similarity 87.0%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAGAGCGCAGACTGGTCTTC 24  
DB 334 AGAAGAGCGTAGACTGGTATTC 356

RESULT 37  
US-09-728-445-172  
; Sequence 172, Application US/09728445  
; Patent No. US20020102543A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: NO. US20020102543A1el Mutated Mammalian Cells and  
; FILE REFERENCE: LEX-0102-USA  
; CURRENT APPLICATION NUMBER: US/09/728,445  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,358  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 891  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172  
; LENGTH: 331  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(331)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-728-445-172

Query Match 71.2%; Score 17.8; DB 3; Length 331;  
Best Local Similarity 90.5%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
DB 184 AAGAGCACAGACTGCTCTTCC 204

RESULT 38  
US-10-964-549-172  
; Sequence 172, Application US/10964549  
; Publication No. US20050186677A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: Novel Mutated Mammalian Cells and Animals  
; FILE REFERENCE: LEX-0286-USA  
; CURRENT APPLICATION NUMBER: US/10/964,549  
; CURRENT FILING DATE: 2004-10-13  
; PRIOR APPLICATION NUMBER: US/09/750,456  
; PRIOR FILING DATE: 2000-12-28

; PRIOR APPLICATION NUMBER: US 09/728,445  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,358  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 891  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172  
; LENGTH: 331  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(331)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-964-549-172

Query Match 71.2%; Score 17.8; DB 9; Length 331;  
Best Local Similarity 90.5%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
DB 184 AAGAGCACAGACTGCTCTTCC 204

RESULT 39  
US-10-359-050-17  
; Sequence 17, Application US/10359050  
; Publication No. US20030186291A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH  
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTEGRASE GENES  
; FILE REFERENCE: AR03-001  
; CURRENT APPLICATION NUMBER: US/10/359,050  
; CURRENT FILING DATE: 2003-02-05  
; PRIOR APPLICATION NUMBER: US 60/354,741  
; PRIOR FILING DATE: 2002-02-06  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 17  
; LENGTH: 4333  
; TYPE: DNA  
; ORGANISM: ROSA 3'  
US-10-359-050-17

Query Match 71.2%; Score 17.8; DB 6; Length 4333;  
Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
DB 1308 AAGAGCACAGACTGCTCTTCC 1328

RESULT 40  
US-10-685-837-2  
; Sequence 2, Application US/10685837  
; Publication No. US20050071893A1  
; GENERAL INFORMATION:  
; APPLICANT: Seibler, Jost  
; APPLICANT: Schwenk, Frieder  
; APPLICANT: Kuhn, Ralf  
; APPLICANT: Kuter-Luks, Birgit  
; TITLE OF INVENTION: siRNA mediated gene silencing in transgenic animals  
; FILE REFERENCE: 022698us JH/BM  
; CURRENT APPLICATION NUMBER: US/10/685,837  
; CURRENT FILING DATE: 2003-10-15  
; PRIOR APPLICATION NUMBER: US60/420,476  
; PRIOR FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: US60/467,814  
; PRIOR FILING DATE: 2003-05-03  
; PRIOR APPLICATION NUMBER: US60/485,969  
; PRIOR FILING DATE: 2003-07-03



```
; PRIOR APPLICATION NUMBER: 02023283.1
; PRIOR FILING DATE: 2002-10-17
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 5409
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: sequence of
; OTHER INFORMATION: homology region
US-10-685-837-2

Query Match          71.2%; Score 17.8; DB 9; Length 5409;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 2384 AAGAGCACAGACTGCTCTTCC 2404

RESULT 41
US-10-359-050-18
; Sequence 18, Application US/10359050
; Publication No. US20030186291A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTEGRASE GENES
; FILE REFERENCE: AR03-001
; CURRENT APPLICATION NUMBER: US/10/359,050
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/354,741
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 10491
; TYPE: DNA
; ORGANISM: PROSA 12
US-10-359-050-18

Query Match          71.2%; Score 17.8; DB 6; Length 10491;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 4954 AAGAGCACAGACTGCTCTTCC 4974

RESULT 42
US-09-738-968-43
; Sequence 43, Application US/09738968
; Patent No. US20010037016A1
; GENERAL INFORMATION:
; APPLICANT: Contag, Pamela R.
; APPLICANT: Purchio, Anthony
; APPLICANT: Zhang, Ning
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING FOR ANGIOGENESIS
; TITLE OF INVENTION: MODULATING COMPOUNDS
; FILE REFERENCE: 9400-0012.20
; CURRENT APPLICATION NUMBER: US/09/738,968
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 09/465,978
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 11176
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-738-968-43
```

```
Query Match          71.2%; Score 17.8; DB 3; Length 11176;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 10526 AAGAGCACAGACTGCTCTTCC 10546

RESULT 43
US-10-359-050-20
; Sequence 20, Application US/10359050
; Publication No. US20030186291A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTEGRASE GENES
; FILE REFERENCE: AR03-001
; CURRENT APPLICATION NUMBER: US/10/359,050
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/354,741
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 11784
; TYPE: DNA
; ORGANISM: C31 substrate reporter
US-10-359-050-20

Query Match          71.2%; Score 17.8; DB 6; Length 11784;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 8760 AAGAGCACAGACTGCTCTTCC 8780

RESULT 44
US-10-014-099F-106
; Sequence 106, Application US/10014099F
; Publication No. US20040003420A1
; GENERAL INFORMATION:
; APPLICANT: KUEHN, Ralf
; APPLICANT: FELDER, Susanne
; APPLICANT: SCHWENK, Frieder
; APPLICANT: KUESTER-LUKS, Birgit
; APPLICANT: FAUST, Nicole
; TITLE OF INVENTION: Modified Recombinase
; FILE REFERENCE: 012787wo/JH/ml
; CURRENT APPLICATION NUMBER: US/10/014,099F
; CURRENT FILING DATE: 2001-11-12
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 11784
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: modified
; OTHER INFORMATION: ROSA26 locus
US-10-014-099F-106

Query Match          71.2%; Score 17.8; DB 6; Length 11784;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 8760 AAGAGCACAGACTGCTCTTCC 8780
```

```
RESULT 45
US-10-359-050-12
; Sequence 12, Application US/10359050
; Publication No. US20030186291A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTEGRASE GENES
; FILE REFERENCE: AR03-001
; CURRENT APPLICATION NUMBER: US/10/359,050
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/354,741
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 12538
; TYPE: DNA
; ORGANISM: PROSA-SA-C31-Int (CNLS)
US-10-359-050-12

Query Match          71.2%; Score 17.8; DB 6; Length 12538;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 AAGAGCGCAGACTGGTCTTCC 25
Db      7001 AAGAGCACAGACTGCTCTTCC 7021

RESULT 46
US-10-789-465-1
; Sequence 1, Application US/10789465
; Publication No. US20040197317A1
; GENERAL INFORMATION:
; APPLICANT: Rao, Mahendra S
; TITLE OF INVENTION: Persistent Expression of Candidate Molecule in Proliferating Stem
; FILE REFERENCE: 2923-5456.1PC
; CURRENT APPLICATION NUMBER: US/10/789,465
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/440,152
; PRIOR FILING DATE: 2003-01-13
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 12607
; TYPE: DNA
; ORGANISM: Mus musculus, synthetic
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2166)
; OTHER INFORMATION: Region: pUC vector
; FEATURE:
; NAME/KEY: gene
; LOCATION: (2191)..(6098)
; OTHER INFORMATION: Polr2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2291)..(2477)
; OTHER INFORMATION: Exon 26 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2817)..(2966)
; OTHER INFORMATION: Exon 27 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3050)..(4206)
; OTHER INFORMATION: Exon 28 of polr2a gene
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (4207)..(6098)
; OTHER INFORMATION: gene=Polr2a

RESULT 47
US-10-867-628-1
; Sequence 1, Application US/10867628
; Publication No. US20050048041A1
; GENERAL INFORMATION:
; APPLICANT: Rao, Mahendra S
; TITLE OF INVENTION: Persistent Expression of Candidate Molecule in Proliferating Stem
; FILE REFERENCE: 2923-5456.1PC
; CURRENT APPLICATION NUMBER: US/10/867,628
; CURRENT FILING DATE: 2004-06-15
; PRIOR APPLICATION NUMBER: US 60/440,152
; PRIOR FILING DATE: 2003-01-13
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 12607
; TYPE: DNA
; ORGANISM: Mus musculus, synthetic
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2166)
; OTHER INFORMATION: Region: pUC vector
; FEATURE:
; NAME/KEY: gene
; LOCATION: (2191)..(6098)
; OTHER INFORMATION: Polr2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2291)..(2477)
; OTHER INFORMATION: Exon 26 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2817)..(2966)
; OTHER INFORMATION: Exon 27 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3050)..(4206)
; OTHER INFORMATION: Exon 28 of polr2a gene
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (4207)..(6098)
; OTHER INFORMATION: gene=Polr2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4469)..(5056)
; OTHER INFORMATION: Region: Internal Ribosome Entry Site (IRES)
; FEATURE:
; NAME/KEY: gene
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4469)..(5056)
; OTHER INFORMATION: Region: Internal Ribosome Entry Site (IRES)
; FEATURE:
; NAME/KEY: gene
; LOCATION: (5057)..(5860)
; OTHER INFORMATION: gene=neo
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12140)..(12607)
; OTHER INFORMATION: Region: pUC vector
US-10-789-465-1

Query Match          71.2%; Score 17.8; DB 8; Length 12607;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 AAGAGCGCAGACTGGTCTTCC 25
Db      11670 AAGAGCGCGACTGCTCTTCC 11690

RESULT 47
US-10-867-628-1
; Sequence 1, Application US/10867628
; Publication No. US20050048041A1
; GENERAL INFORMATION:
; APPLICANT: Rao, Mahendra S
; TITLE OF INVENTION: Persistent Expression of Candidate Molecule in Proliferating Stem
; FILE REFERENCE: 2923-5456.1PC
; CURRENT APPLICATION NUMBER: US/10/867,628
; CURRENT FILING DATE: 2004-06-15
; PRIOR APPLICATION NUMBER: US 60/440,152
; PRIOR FILING DATE: 2003-01-13
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 12607
; TYPE: DNA
; ORGANISM: Mus musculus, synthetic
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2166)
; OTHER INFORMATION: Region: pUC vector
; FEATURE:
; NAME/KEY: gene
; LOCATION: (2191)..(6098)
; OTHER INFORMATION: Polr2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2291)..(2477)
; OTHER INFORMATION: Exon 26 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2817)..(2966)
; OTHER INFORMATION: Exon 27 of polr2a gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3050)..(4206)
; OTHER INFORMATION: Exon 28 of polr2a gene
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (4207)..(6098)
; OTHER INFORMATION: gene=Polr2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4469)..(5056)
; OTHER INFORMATION: Region: Internal Ribosome Entry Site (IRES)
; FEATURE:
; NAME/KEY: gene
```

```
; LOCATION: (5057)..(5860)
; OTHER INFORMATION: gene=neo
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12140)..(12607)
; OTHER INFORMATION: Region: pUC vector
US-10-867-628-1

Query Match          71.2%; Score 17.8; DB 8; Length 12607;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 11670 AAGAGCGCAGACTGGTCTTCC 11690

RESULT 48
US-10-359-050-13
; Sequence 13, Application US/10359050
; Publication No. US20030186291A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS DEUTSCHLAND GMBH
; TITLE OF INVENTION: GENETICALLY ENGINEERED PHIC31-INTERGASE GENES
; FILE REFERENCE: AR03-001
; CURRENT APPLICATION NUMBER: US/10/359,050
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/354,741
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 12645
; TYPE: DNA
; ORGANISM: PROSA-SA-C31-Int (CNLS) -CO
US-10-359-050-13

Query Match          71.2%; Score 17.8; DB 6; Length 12645;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 7108 AAGAGCGCAGACTGGTCTTCC 7128

RESULT 49
US-10-685-837-1
; Sequence 1, Application US/10685837
; Publication No. US20050071893A1
; GENERAL INFORMATION:
; APPLICANT: Seibler, Jost
; APPLICANT: Schwenk, Frieder
; APPLICANT: 'Kuhn, Ralf
; APPLICANT: Kuter-Luks, Birgit
; TITLE OF INVENTION: siRNA mediated gene silencing in transgenic animals
; FILE REFERENCE: 022698us JH/BM
; CURRENT APPLICATION NUMBER: US/10/685,837
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US60/420,476
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: US60/467,814
; PRIOR FILING DATE: 2003-05-03
; PRIOR APPLICATION NUMBER: US60/485,969
; PRIOR FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: 02023283.1
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 13139
; TYPE: DNA
; ORGANISM: Artificial Sequence

; LOCATION: (5057)..(5860)
; OTHER INFORMATION: Description of Artificial Sequence: Rosa26 locus
; OTHER INFORMATION: sequence
US-10-685-837-1

Query Match          71.2%; Score 17.8; DB 9; Length 13139;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 4940 AAGAGCACAGACTGCTCTTCC 4960

RESULT 50
US-10-685-837-4
; Sequence 4, Application US/10685837
; Publication No. US20050071893A1
; GENERAL INFORMATION:
; APPLICANT: Seibler, Jost
; APPLICANT: Schwenk, Frieder
; APPLICANT: Kuhn, Ralf
; APPLICANT: Kuter-Luks, Birgit
; TITLE OF INVENTION: siRNA mediated gene silencing in transgenic animals
; FILE REFERENCE: 022698us JH/BM
; CURRENT APPLICATION NUMBER: US/10/685,837
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US60/420,476
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: US60/467,814
; PRIOR FILING DATE: 2003-05-03
; PRIOR APPLICATION NUMBER: US60/485,969
; PRIOR FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: 02023283.1
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 14947
; TYPE: DNA
; ORGANISM: Artificial Sequence

; LOCATION: (5057)..(5860)
; OTHER INFORMATION: Description of Artificial Sequence: Targeting
; OTHER INFORMATION: vector for Rosa26 locus with a Fluc-hygro insert
US-10-685-837-4

Query Match          71.2%; Score 17.8; DB 9; Length 14947;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
Db 6824 AAGAGCACAGACTGCTCTTCC 6844

Search completed: February 3, 2006, 15:44:12
Job time : 372.556 secs
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2006, 11:35:31 ; Search time 343.111 Seconds  
(without alignments)  
61.064 Million cell updates/sec

Title: US-10-719-900-35

Perfect score: 25

Sequence: 1 aaaaagagcgagactgtcttc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6068529 seqs, 419036697 residues

Total number of hits satisfying chosen parameters: 12137058

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications NA New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	75.2	43445	8	US-11-124-020A-1
C 2	18.8	75.2	43445	8	US-11-124-020A-2
C 3	18.2	72.8	3161	7	US-10-750-185-32958
C 4	18.2	72.8	3161	7	US-10-750-623-32958
5	18.2	72.8	7715	8	US-11-147-109-1
6	17.8	71.2	95	7	US-10-909-125-1605
7	17.8	71.2	2576	8	US-11-136-527-3672
8	17.8	71.2	2712	8	US-11-136-527-2401
9	17.8	71.2	171936	7	US-10-933-025-24
C 10	17.6	70.4	580	8	US-11-128-061-1155
C 11	17.6	70.4	580	8	US-11-128-061-4797
C 12	17.6	70.4	580	8	US-11-128-049-1155
C 13	17.6	70.4	580	8	US-11-128-049-4797
C 14	17.6	70.4	2180	7	US-10-750-185-48791
C 15	17.6	70.4	2180	7	US-10-750-623-48791
C 16	17.6	70.4	2298	8	US-11-037-243-21
17	17.2	68.8	1574	7	US-10-750-185-33387
18	17.2	68.8	1574	7	US-10-750-623-33387
C 19	17	68.0	201	7	US-10-995-561-76336
C 20	17	68.0	1630	8	US-11-136-527-711
21	17	68.0	2576	7	US-10-750-185-62716
22	17	68.0	2576	7	US-10-750-623-62716

C 23	68.0	2929	7	US-10-750-185-43029	Sequence 43029, A
C 24	68.0	2929	7	US-10-750-623-43029	Sequence 43029, A
C 25	68.0	394468	7	US-10-995-561-13473	Sequence 13473, A
C 26	67.2	481	8	US-11-000-688-1233	Sequence 1233, Ap
C 27	67.2	1897	7	US-10-750-185-45468	Sequence 45468, A
C 28	67.2	1897	7	US-10-750-623-45468	Sequence 45468, A
C 29	66.4	578	8	US-11-128-061-1621	Sequence 1621, Ap
C 30	66.4	578	8	US-11-128-061-5263	Sequence 5263, Ap
C 31	66.4	578	8	US-11-128-049-1621	Sequence 1621, Ap
C 32	66.4	578	8	US-11-128-049-5263	Sequence 5263, Ap
C 33	66.4	157224	8	US-11-112-908-51	Sequence 51, Appl
C 34	66.4	170189	8	US-11-112-908-51	Sequence 51, Appl
C 35	65.6	3684	7	US-10-750-185-51696	Sequence 51696, A
C 36	65.6	3684	7	US-10-750-623-51696	Sequence 51696, A
C 37	64.8	21	7	US-10-770-726-21824	Sequence 21824, A
C 38	64.8	196	7	US-10-502-972-25	Sequence 25, Appl
C 39	64.8	201	7	US-10-995-561-32523	Sequence 32523, A
C 40	64.8	201	7	US-10-995-561-57056	Sequence 57056, A
C 41	64.8	201	7	US-10-995-561-65863	Sequence 65863, A
C 42	64.8	275	8	US-11-128-061-3322	Sequence 3322, Ap
C 43	64.8	275	8	US-11-128-061-6964	Sequence 6964, Ap
C 44	64.8	275	8	US-11-128-049-3322	Sequence 3322, Ap
C 45	64.8	275	8	US-11-128-049-6964	Sequence 6964, Ap
C 46	64.8	425	8	US-11-128-061-4682	Sequence 4682, Ap
C 47	64.8	425	8	US-11-128-049-4682	Sequence 4682, Ap
C 48	64.8	431	8	US-11-108-172-584	Sequence 584, App
C 49	64.8	469	8	US-11-128-061-1040	Sequence 1040, Ap
C 50	64.8	469	8	US-11-128-049-1040	Sequence 1040, Ap
C 51	64.8	538	8	US-11-128-061-2708	Sequence 2708, Ap
C 52	64.8	538	8	US-11-128-061-6350	Sequence 6350, Ap
C 53	64.8	538	8	US-11-128-049-2708	Sequence 2708, Ap
C 54	64.8	538	8	US-11-128-049-6350	Sequence 6350, Ap
C 55	64.8	581	8	US-11-128-061-1928	Sequence 1928, Ap
C 56	64.8	581	8	US-11-128-061-5570	Sequence 5570, Ap
C 57	64.8	581	8	US-11-128-049-1928	Sequence 1928, Ap
C 58	64.8	581	8	US-11-128-049-5570	Sequence 5570, Ap
C 59	64.8	600	8	US-11-136-527-6661	Sequence 6661, Ap
C 60	64.8	600	8	US-11-136-527-7498	Sequence 7498, Ap
C 61	64.8	600	8	US-11-136-527-7834	Sequence 7834, Ap
C 62	64.8	600	8	US-11-136-527-7913	Sequence 7913, Ap
C 63	64.8	600	8	US-11-136-527-8182	Sequence 8182, Ap
C 64	64.8	654	8	US-11-128-061-710	Sequence 710, App
C 65	64.8	654	8	US-11-128-049-710	Sequence 710, App
C 66	64.8	1201	8	US-11-136-527-3817	Sequence 3817, Ap
C 67	64.8	1400	8	US-11-136-527-7985	Sequence 7985, Ap
C 68	64.8	1631	8	US-11-128-061-3387	Sequence 3387, Ap
C 69	64.8	1631	8	US-11-128-049-3387	Sequence 3387, Ap
C 70	64.8	1659	8	US-11-136-527-2565	Sequence 2565, Ap
C 71	64.8	2040	8	US-11-136-527-3665	Sequence 3665, Ap
C 72	64.8	2228	8	US-11-136-527-4091	Sequence 4091, Ap
C 73	64.8	2509	8	US-11-136-527-3738	Sequence 3738, Ap
C 74	64.8	2684	8	US-11-136-527-3527	Sequence 3527, Ap
C 75	64.8	2893	8	US-11-136-527-3005	Sequence 3005, Ap
C 76	64.8	2820	7	US-10-770-726-20	Sequence 20, Appl
C 77	64.8	3262	8	US-11-136-527-3402	Sequence 3402, Ap
C 78	64.8	3652	8	US-11-136-527-1929	Sequence 1929, Ap
C 79	64.8	4417	8	US-11-136-527-606	Sequence 606, App
C 80	64.8	5762	7	US-11-136-527-4086	Sequence 4086, Ap
C 81	64.8	8196	7	US-10-678-790-53	Sequence 53, Appl
C 82	64.8	8567	8	US-11-128-061-3390	Sequence 3390, Ap
C 83	64.8	8567	8	US-11-128-049-3390	Sequence 3390, Ap
C 84	64.8	9196	7	US-10-959-310-13	Sequence 13, Appl
C 85	64.8	9196	8	US-11-131-212-3	Sequence 3, Appl
C 86	64.8	9578	8	US-11-136-527-3985	Sequence 3985, Ap
C 87	64.8	13187	8	US-11-136-527-3585	Sequence 3585, Ap
C 88	64.8	13206	8	US-11-136-527-4042	Sequence 4042, Ap
C 89	64.8	13894	8	US-11-136-527-3889	Sequence 3889, Ap
C 90	64.8	14465	8	US-11-128-061-728	Sequence 728, App
C 91	64.8	14465	8	US-11-128-049-728	Sequence 728, App
C 92	64.8	14804	7	US-10-995-561-13379	Sequence 13379, A
C 93	64.8	15165	8	US-11-136-527-3897	Sequence 3897, Ap
C 94	64.8	15165	8	US-11-136-527-3897	Sequence 3897, Ap
C 95	64.8	16257	8	US-11-136-527-4054	Sequence 4054, Ap

```

96 16.2 64.8 17517 8 US-11-136-527-3650
97 16.2 64.8 17517 8 US-11-136-527-3650
98 16.2 64.8 17517 8 US-11-136-527-3650
99 16.2 64.8 17517 8 US-11-136-527-3650
100 16.2 64.8 17517 8 US-11-136-527-3650
101 16.2 64.8 17517 8 US-11-136-527-3650
102 16.2 64.8 17517 8 US-11-136-527-3650
103 16.2 64.8 17517 8 US-11-136-527-3650
104 16.2 64.8 17517 8 US-11-136-527-3650
105 16.2 64.8 17517 8 US-11-136-527-3650
106 16.2 64.8 17517 8 US-11-136-527-3650
107 16.2 64.8 17517 8 US-11-136-527-3650
108 16.2 64.8 17517 8 US-11-136-527-3650
109 16.2 64.8 17517 8 US-11-136-527-3650
110 16.2 64.8 17517 8 US-11-136-527-3650
111 16.2 64.8 17517 8 US-11-136-527-3650
112 16.2 64.8 17517 8 US-11-136-527-3650
113 16.2 64.8 17517 8 US-11-136-527-3650
114 16.2 64.8 17517 8 US-11-136-527-3650
115 16.2 64.8 17517 8 US-11-136-527-3650
116 16.2 64.8 17517 8 US-11-136-527-3650
117 16.2 64.8 17517 8 US-11-136-527-3650
118 16.2 64.8 17517 8 US-11-136-527-3650
119 16.2 64.8 17517 8 US-11-136-527-3650
120 16.2 64.8 17517 8 US-11-136-527-3650
121 16.2 64.8 17517 8 US-11-136-527-3650
122 16.2 64.8 17517 8 US-11-136-527-3650
123 16.2 64.8 17517 8 US-11-136-527-3650
124 16.2 64.8 17517 8 US-11-136-527-3650
125 16.2 64.8 17517 8 US-11-136-527-3650
126 16.2 64.8 17517 8 US-11-136-527-3650
127 16.2 64.8 17517 8 US-11-136-527-3650
128 16.2 64.8 17517 8 US-11-136-527-3650
129 16.2 64.8 17517 8 US-11-136-527-3650
130 16.2 64.8 17517 8 US-11-136-527-3650
131 16.2 64.8 17517 8 US-11-136-527-3650
132 16.2 64.8 17517 8 US-11-136-527-3650
133 15.8 63.2 201 7 US-10-995-561-32198
134 15.8 63.2 600 7 US-10-750-185-32198
135 15.8 63.2 600 7 US-10-750-185-32198
136 15.8 63.2 1109 7 US-10-750-623-21237
137 15.8 63.2 1109 7 US-10-750-185-35454
138 15.8 63.2 2196 7 US-10-750-623-35454
139 15.8 63.2 2196 7 US-10-750-185-44622
140 15.8 63.2 2196 7 US-10-750-623-44622
141 15.8 63.2 2452 8 US-11-136-527-464
142 15.8 63.2 131855 8 US-11-112-908-29
143 15.8 63.2 150173 8 US-11-112-908-26
144 15.8 63.2 166020 8 US-11-112-908-28
145 15.8 63.2 171247 8 US-11-112-908-27
146 15.8 63.2 187745 8 US-11-121-086-83
147 15.8 63.2 387780 7 US-10-995-561-13259
148 15.6 62.4 25 8 US-11-121-849-490854
149 15.6 62.4 201 7 US-10-995-561-44151
150 15.6 62.4 583 8 US-11-128-061-2491
```

ALIGNMENTS

```

RESULT 1
US-11-124-020A-1
; Sequence 1, Application US/11124020A
; Publication No. US20050287558A1
; GENERAL INFORMATION:
; APPLICANT: Rosanne M. Crooke
; APPLICANT: Steven Mah
; APPLICANT: Steven Mah
; TITLE OF INVENTION: SNPS OF APOLIPOPROTEIN B AND MODULATION
; FILE REFERENCE: BIO0021US
; CURRENT APPLICATION NUMBER: US/11/124,020A
; CURRENT FILING DATE: 2005-05-05

US-10-750-185-32958/c
; Sequence 32958, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32958
; LENGTH: 3161
```

```

; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 43445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-11-124-020A-1

Query Match 75.2%; Score 18.8; DB 8; Length 43445;
Best Local Similarity 90.9%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCT 22
Db 25768 AAAACAGAGCGTAGACTGGTCT 25789

RESULT 2
US-11-124-020A-2/c
; Sequence 2, Application US/11124020A
; Publication No. US20050287558A1
; GENERAL INFORMATION:
; APPLICANT: Rosanne M. Crooke
; APPLICANT: Mark J. Graham
; APPLICANT: Steven Mah
; TITLE OF INVENTION: SNPS OF APOLIPOPROTEIN B AND MODULATION
; FILE REFERENCE: BIO0021US
; CURRENT APPLICATION NUMBER: US/11/124,020A
; CURRENT FILING DATE: 2005-05-05
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 43445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-11-124-020A-2

Query Match 75.2%; Score 18.8; DB 8; Length 43445;
Best Local Similarity 90.9%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGGTCT 22
Db 17678 AAAACAGAGCGTAGACTGGTCT 17657

RESULT 3
US-10-750-185-32958/c
; Sequence 32958, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32958
; LENGTH: 3161
```

Query Match 72.8%; Score 18.2; DB 8; Length 7715;

```

/ CURRENT FILING DATE: 2005-05-25
/ PRIOR APPLICATION NUMBER: US 60/574,294
/
/ PRIOR FILING DATE: 2005-05-26
/
/ NUMBER OF SEQ ID NOS: 362830
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 3672
/ LENGTH: 2576
/

```



```
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1147)..(1147)
; OTHER INFORMATION: n is a, c, g, or t
US-11-136-527-3672

Query Match          71.2%; Score 17.8; DB 8; Length 2576;
Best Local Similarity 90.5%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| |||||
Db 814 AAGAGCACAGACTGCTCTTCC 834

RESULT 8
US-11-136-527-2401
; Sequence 2401, Application US/111136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Mounts, William M
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2401
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2401

Query Match          71.2%; Score 17.8; DB 8; Length 2712;
Best Local Similarity 90.5%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| |||||
Db 87 AAGAGCAGTGGTCTTCC 107

RESULT 9
US-10-933-025-24
; Sequence 24, Application US/10933025
; Publication No. US20050265967A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN, STEVEN
; APPLICANT: HEMMERICH, STEFAN
; APPLICANT: TOMITA, MEGUMI
; TITLE OF INVENTION: Sulfotransferases and methods of use
; FILE REFERENCE: UCAL-230CON
; CURRENT APPLICATION NUMBER: US/10/933,025
; CURRENT FILING DATE: 2004-09-01
; PRIOR APPLICATION NUMBER: 10/025,966
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/258,577
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 60/267,831
; PRIOR FILING DATE: 2001-09-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 171936
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-933-025-24

; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1147)..(1147)
; OTHER INFORMATION: n is a, c, g, or t
US-11-136-527-3672

Query Match          71.2%; Score 17.8; DB 7; Length 171936;
Best Local Similarity 90.5%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
    ||||| ||||| ||||| |||||
Db 144637 AAGAGCACAGACTGCTCTTCC 144657

RESULT 10
US-11-128-061-1155/c
; Sequence 1155, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1155
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Cricetulus griseus
US-11-128-061-1155

Query Match          70.4%; Score 17.6; DB 8; Length 580;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTGGTCTTC 24
    ||||| ||||| ||||| |||||
Db 53 AAAAAGAGCACAGACTAGTCTTC 30

RESULT 11
US-11-128-061-4797/c
; Sequence 4797, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4797
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Cricetulus griseus
US-11-128-061-4797
```

US-11-128-061-4797

Query Match 70.4%; Score 17.6; DB 8; Length 580;  
Best Local Similarity 83.3%; Pred. No. 47;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 24  
||||| ||| ||||| ||||| |||||  
DB 53 AAAAAACAGCAGACTAGTCTTC 30

RESULT 12

US-11-128-049-1155/c  
; Sequence 1155, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 1155  
; LENGTH: 580  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-049-1155

Query Match 70.4%; Score 17.6; DB 8; Length 580;  
Best Local Similarity 83.3%; Pred. No. 47;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 24  
||||| ||| ||||| ||||| |||||  
DB 53 AAAAAACAGCAGACTAGTCTTC 30

RESULT 13

US-11-128-049-4797/c  
; Sequence 4797, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 4797  
; LENGTH: 580  
; TYPE: DNA

; ORGANISM: Cricetulus griseus  
US-11-128-049-4797

Query Match 70.4%; Score 17.6; DB 8; Length 580;  
Best Local Similarity 83.3%; Pred. No. 47;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 24  
||||| ||| ||||| ||||| |||||  
DB 53 AAAAAACAGCAGACTAGTCTTC 30

RESULT 14

US-10-750-185-48791/c  
; Sequence 48791, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 48791  
; LENGTH: 2180  
; TYPE: DNA  
; ORGANISM: Bovine 19866881276311  
US-10-750-185-48791

Query Match 70.4%; Score 17.6; DB 7; Length 2180;  
Best Local Similarity 83.3%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAGCGCAGACTGGTCTTC 24  
||||| ||| ||||| ||||| |||||  
DB 1120 AGAACAGAGTACAGACTGGTCTTC 1097

RESULT 15

US-10-750-623-48791/c  
; Sequence 48791, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 48791  
; LENGTH: 2180  
; TYPE: DNA  
; ORGANISM: Bovine 19866881276311  
US-10-750-623-48791



```
; Sequence 711, Application US/11136527
; Publication No. US2005028750A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 711
; LENGTH: 1630
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-711

Query Match      68.0%; Score 17; DB 8; Length 1630;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAGAGCGCAGACTGGTCTTCC 25
      ||||| ||||| ||||| ||||| |||||
Db      159 AAGAAATAGCTCAACGGGTCTTCC 135

RESULT 21
US-10-750-185-62716
; Sequence 62716, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62716
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Bovine 19866880397212
US-10-750-185-62716

Query Match      68.0%; Score 17; DB 7; Length 2576;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAGAGCGCAGACTGGTCTTCC 25
      ||||| ||||| ||||| ||||| |||||
Db      2387 AAAAAAGTGCAGATACGTCTTCC 2411

RESULT 22
US-10-750-623-62716
; Sequence 62716, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62716
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Bovine 19866880397212
US-10-750-623-62716

Query Match      68.0%; Score 17; DB 7; Length 2576;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAGAGCGCAGACTGGTCTTCC 25
      ||||| ||||| ||||| ||||| |||||
Db      652 AAAAAAGCAAGAGTGATCTTCC 628

RESULT 24
US-10-750-623-43029/c
; Sequence 43029, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43029
; LENGTH: 2929
; TYPE: DNA
; ORGANISM: Bovine 19866880892057
US-10-750-185-43029

Query Match      68.0%; Score 17; DB 7; Length 2929;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAGAGCGCAGACTGGTCTTCC 25
      ||||| ||||| ||||| ||||| |||||
Db      652 AAAAAAGCAAGAGTGATCTTCC 628

RESULT 23
US-10-750-185-43029/c
; Sequence 43029, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43029
; LENGTH: 2929
; TYPE: DNA
; ORGANISM: Bovine 19866880892057
US-10-750-185-43029

Query Match      68.0%; Score 17; DB 7; Length 2929;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 AAAAAAGAGCGCAGACTGGTCTTCC 25
      ||||| ||||| ||||| ||||| |||||
Db      652 AAAAAAGCAAGAGTGATCTTCC 628

RESULT 24
US-10-750-623-43029/c
; Sequence 43029, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
```

```
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43029
; LENGTH: 2929
; TYPE: DNA
; ORGANISM: Bovine 19866880892057
US-10-750-623-43029

Query Match      68.0%; Score 17; DB 7; Length 2929;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTCC 25
   ||||| ||||| ||||| ||||| |||||
Db 652 AAAAAAGAGCCAGAGTGATCTCC 628

RESULT 25
US-10-995-561-13473/c
; Sequence 13473, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13473
; LENGTH: 394468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..-(394468)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13473

Query Match      68.0%; Score 17; DB 7; Length 394468;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTGCTCTCC 25
   ||||| ||||| ||||| ||||| |||||
Db 327726 AACAAAGAGCGCAGACTCCACTCC 327702

RESULT 26
US-11-000-688-1233
; Sequence 1233, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, Francois
; APPLICANT: HOULGATTE, Remi
; APPLICANT: BIRNBAUM, Daniel
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
; FILE REFERENCE: 1423-R-03
; CURRENT APPLICATION NUMBER: US/11/000,688
; CURRENT FILING DATE: 2004-12-01
; PRIOR APPLICATION NUMBER: US 60/525,987
; PRIOR FILING DATE: 2003-12-01
; NUMBER OF SEQ ID NOS: 1596
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1233
; LENGTH: 481
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequences:primer
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(481)
; OTHER INFORMATION: 3' terminal sequence from clone
; OTHER INFORMATION: image:512924.
US-11-000-688-1233

Query Match      67.2%; Score 16.8; DB 8; Length 481;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAGCGCAGACTGGTC 21
   ||||| ||||| ||||| ||||| |||||
Db 85 AAAAGGAGCAGACTGGTC 104

RESULT 27
US-10-750-185-45468/c
; Sequence 45468, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45468
; LENGTH: 1897
; TYPE: DNA
; ORGANISM: Bovine 19866880485767
US-10-750-185-45468

Query Match      67.2%; Score 16.8; DB 7; Length 1897;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAGAGCGCAGACTGGTCTT 23
   ||||| ||||| ||||| ||||| |||||
Db 900 AAAAAGTGCAGACTGGTCTT 881

RESULT 28
US-10-750-623-45468/c
; Sequence 45468, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```



```
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5263
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Cricetus griseus
US-11-128-049-5263

Query Match      66.4%; Score 16.6; DB 8; Length 578;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAGAGCGCAGACTGGTCTTC 25
   ||||| ||||| ||||| |||||
Db 66 AAGAGGGTGCAGATTGGTCTTC 44

RESULT 33
US-11-112-908-51
; Sequence 51, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 51
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-51

Query Match      66.4%; Score 16.6; DB 8; Length 157224;
Best Local Similarity 82.6%; Pred. No. 3.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTGGTCTT 23
   ||||| | ||||| ||||| |||||
Db 75474 AAAAAAATCGCAGACTGCTT 75496

RESULT 34
US-11-112-908-50
; Sequence 50, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
```

---

```
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 50
; LENGTH: 170189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-50

Query Match      66.4%; Score 16.6; DB 8; Length 170189;
Best Local Similarity 82.6%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTGGTCTT 23
   ||||| ||||| ||||| ||||| |||||
Db 108234 AAAAAAATCGCAGACTGCTT 108256

RESULT 35
US-10-750-185-51696/c
; Sequence 51696, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51696
; LENGTH: 3684
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-51696

Query Match      65.6%; Score 16.4; DB 7; Length 3684;
Best Local Similarity 94.4%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAGAGCGCAGACTG 18
   ||||| ||||| ||||| |||||
Db 2655 AAAAAAGAGCAGACTG 2638

RESULT 36
US-10-750-623-51696/c
; Sequence 51696, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
```



```
; SEQ ID NO 51696
; LENGTH: 3684
; TYPE: DNA
; ORGANISM: Bovine 1986680775622
US-10-750-623-51696

Query Match          65.6%; Score 16.4; DB 7; Length 3684;
Best Local Similarity 94.4%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAGAGCGCAGACTG 18
    ||||| ||||| ||||| |||||
Db 2655 AAAAAAGAGCAGACTG 2638

RESULT 37
US-10-770-726-21824
; Sequence 21824, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Brown, Eugene
; APPLICANT: Liu, Wei
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: AM101079 (031896-010000)
; CURRENT APPLICATION NUMBER: US/10/770,726
; CURRENT FILING DATE: 2004-02-04
; NUMBER OF SEQ ID NOS: 48640
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21824
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-770-726-21824

Query Match          64.8%; Score 16.2; DB 7; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAGCGCAGACTGCTTT 23
    ||||| ||||| ||||| |||||
Db 1 AATGAGAGGAGCTGCTTT 21

RESULT 38
US-10-502-972-25
; Sequence 25, Application US/10502972
; Publication No. US2005025573A1
; GENERAL INFORMATION:
; APPLICANT: Chambers, Ian
; APPLICANT: Smith, Austin G.
; TITLE OF INVENTION: Pluripotency determining factors and uses thereof
; FILE REFERENCE: 2325.002000/RVE/RAS
; CURRENT APPLICATION NUMBER: US/10/502,972
; CURRENT FILING DATE: 2004-07-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 196
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-502-972-25

Query Match          64.8%; Score 16.2; DB 7; Length 196;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTTCC 25
    ||||| ||||| ||||| |||||
Db 27 AAGAGCAGCTGCTCTTCC 47

RESULT 39
US-10-995-561-32523/c
; Sequence 32523, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32523
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-32523

Query Match          64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AAAGAGCGCAGACTGCTTTC 24
    ||||| ||||| ||||| |||||
Db 140 AAAGAGGAGGAGCTGGCCTTC 120

RESULT 40
US-10-995-561-57056/c
; Sequence 57056, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57056
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-57056

Query Match          64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGAGCGCAGACTGCTTTC 23
    ||||| ||||| ||||| |||||
Db 149 AATGAGAGGAGCTGCTTTC 129

RESULT 41
US-10-995-561-65863
; Sequence 65863, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65863
```

```
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-65863

Query Match          64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAGAGCGCAGACTGGTCTTT 23
   ||||| ||||| ||||| |||||
Db 53 AAATGAGAGGAGCTGGTCTT 73

RESULT 42
US-11-128-061-3322
; Sequence 3322, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3322
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Cricetulus longicaudatus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (172)..(192)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-3322

Query Match          64.8%; Score 16.2; DB 8; Length 275;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
   ||||| ||||| ||||| |||||
Db 198 AAGAGCACTGACTGCTCTTCC 218

RESULT 43
US-11-128-061-6964
; Sequence 6964, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
```

```
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6964
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Cricetulus longicaudatus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (172)..(192)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-6964

Query Match          64.8%; Score 16.2; DB 8; Length 275;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
   ||||| ||||| ||||| |||||
Db 198 AAGAGCACTGACTGCTCTTCC 218

RESULT 44
US-11-128-049-3322
; Sequence 3322, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3322
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Cricetulus longicaudatus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (172)..(192)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-3322

Query Match          64.8%; Score 16.2; DB 8; Length 275;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25
   ||||| ||||| ||||| |||||
Db 198 AAGAGCACTGACTGCTCTTCC 218

RESULT 45
US-11-128-049-6964
; Sequence 6964, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
```

APPLICANT: Sinacore, Martin S.  
APPLICANT: Leonard, Mark W.  
APPLICANT: Brown, Eugene L.  
APPLICANT: Miller, Christopher P.  
TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR  
FILE REFERENCE: 01997.027700  
CURRENT APPLICATION NUMBER: US/11/128,049  
CURRENT FILING DATE: 2005-05-11  
PRIOR APPLICATION NUMBER: US 60/570,425  
PRIOR FILING DATE: 2004-05-11  
NUMBER OF SEQ ID NOS: 7285  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 6964  
LENGTH: 275  
TYPE: DNA  
ORGANISM: Cricetulus longicaudatus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (172)..(192)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-128-049-6964

Query Match 64.8%; Score 16.2; DB 8; Length 275;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
||||| ||||| ||||| |||||  
Db 198 AAGAGCACTGACTGCTCTTCC 218

RESULT 46  
US-11-128-061-4682  
Sequence 4682, Application US/11128061  
Publication No. US20060003958A1  
GENERAL INFORMATION:  
APPLICANT: Melville, Mark W.  
APPLICANT: Charlebois, Timothy S.  
APPLICANT: Mounts, William M.  
APPLICANT: Hann, Louane E.  
APPLICANT: Sinacore, Martin S.  
APPLICANT: Leonard, Mark W.  
APPLICANT: Brown, Eugene L.  
APPLICANT: Miller, Christopher P.  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS  
FILE REFERENCE: 01997.027701  
CURRENT APPLICATION NUMBER: US/11/128,061  
CURRENT FILING DATE: 2005-05-11  
PRIOR APPLICATION NUMBER: US 60/570,425  
PRIOR FILING DATE: 2004-05-11  
NUMBER OF SEQ ID NOS: 7285  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 4682  
LENGTH: 425  
TYPE: DNA  
ORGANISM: Cricetulus griseus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (370)..(390)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-128-061-4682

Query Match 64.8%; Score 16.2; DB 8; Length 425;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
||||| ||||| ||||| |||||  
Db 396 AAGAGCACTGACTGCTCTTCC 416

RESULT 47  
US-11-128-049-4682  
Sequence 4682, Application US/11128049  
Publication No. US20060010513A1  
GENERAL INFORMATION:  
APPLICANT: Melville, Mark W.  
APPLICANT: Charlebois, Timothy S.  
APPLICANT: Mounts, William M.  
APPLICANT: Hann, Louane E.  
APPLICANT: Sinacore, Martin S.  
APPLICANT: Leonard, Mark W.  
APPLICANT: Brown, Eugene L.  
APPLICANT: Miller, Christopher P.  
TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FO  
FILE REFERENCE: 01997.027700  
CURRENT APPLICATION NUMBER: US/11/128,049  
CURRENT FILING DATE: 2005-05-11  
PRIOR APPLICATION NUMBER: US 60/570,425  
PRIOR FILING DATE: 2004-05-11  
NUMBER OF SEQ ID NOS: 7285  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 4682  
LENGTH: 425  
TYPE: DNA  
ORGANISM: Cricetulus griseus  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (370)..(390)  
OTHER INFORMATION: n is a, c, g, or t  
US-11-128-049-4682

Query Match 64.8%; Score 16.2; DB 8; Length 425;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGGTCTTCC 25  
||||| ||||| ||||| |||||  
Db 396 AAGAGCACTGACTGCTCTTCC 416

RESULT 48  
US-11-108-172-584/c  
Sequence 584, Application US/11108172  
Publication No. US20050260177A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Benson, Darin R.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Stolk, John A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Jiang, Yuqiu  
APPLICANT: Smith, Carole L.  
APPLICANT: King, Gordon E.  
APPLICANT: Wang, Aijun  
APPLICANT: Clapper, Jonathan D.  
APPLICANT: Skeiky, Yasir A. W.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick Thomas S.  
APPLICANT: Carter, Darrick  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
FILE REFERENCE: 210121.471C15  
CURRENT APPLICATION NUMBER: US/11/108,172  
CURRENT FILING DATE: 2005-04-15  
PRIOR APPLICATION NUMBER: US 10/025,380  
PRIOR FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: US 09/922,217  
PRIOR FILING DATE: 2001-08-03  
PRIOR APPLICATION NUMBER: US 09/833,263  
PRIOR FILING DATE: 2001-04-10

```
; PRIOR APPLICATION NUMBER: US 09/649,811
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/609,448
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/575,251
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,444
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 09/504,629
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US 09/480,321
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: US 09/476,296
; PRIOR FILING DATE: 1998-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 584
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-108-172-584

Query Match      64.8%; Score 16.2; DB 8; Length 431;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25
    ||||| ||||| |||||
Db 122 AAGAGCACTGACTGTCTTCC 102

RESULT 49
US-11-128-061-1040
; Sequence 1040, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1040
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; NAME/KEY: misc feature
; LOCATION: (370)..(390)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (426)..(450)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (453)..(465)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-1040

Query Match      64.8%; Score 16.2; DB 8; Length 469;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25
    ||||| ||||| |||||
Db 396 AAGAGCACTGACTGTCTTCC 416

RESULT 50
US-11-128-049-1040
; Sequence 1040, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1040
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; NAME/KEY: misc feature
; LOCATION: (370)..(390)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (426)..(450)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (453)..(465)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-1040

Query Match      64.8%; Score 16.2; DB 8; Length 469;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAGAGCGCAGACTGCTCTTCC 25
    ||||| ||||| |||||
Db 396 AAGAGCACTGACTGTCTTCC 416

Search completed: February 3, 2006, 16:20:24
Job time : 346.111 secs
```